

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

Estudio cristalográfico de proteínas implicadas en la ruta de los inositoles polifosfatados y su relación con el ARN

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Elsa Franco Echevarría

DIRECTORAS

Beatriz González Pérez
Juliana Sanz Aparicio

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

**Estudio cristalográfico de proteínas implicadas en la ruta de los
inositoles polifosfatados y su relación con el ARN**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Elsa Franco Echevarría

Directores

Beatriz González Pérez

Juliana Sanz Aparicio

Madrid, 2017

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR I



TESIS DOCTORAL

**Estudio cristalográfico de proteínas implicadas en la ruta de los
inositoles polifosfatados y su relación con el ARN**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Elsa Franco Echevarría

Directores

Beatriz González Pérez

Juliana Sanz Aparicio

DEPARTAMENTO DE CRISTALOGRAFÍA Y BIOLOGÍA ESTRUCTURAL

INSTITUTO DE QUÍMICA FÍSICA “ROCASOLANO”

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA y BIOLOGÍA MOLECULAR I

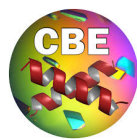


TESIS DOCTORAL

Estudio cristalográfico de proteínas implicadas en la ruta de los inositoles polifosfatados y su relación con el ARN

Memoria presentada por Elsa Franco Echevarría para optar al Grado de Doctor

Esta Tesis Doctoral ha sido realizada en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano” (Consejo superior de Investigaciones Científicas, CSIC) bajo la dirección de la Dra. Beatriz González Pérez (Científica Titular) y la Dra. Juliana Sanz Aparicio (Investigador Científico) ambas miembros del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física “Rocasolano”, CSIC.



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

La Dra. Beatriz González Pérez, Científico Titular del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" del Consejo Superior de Investigaciones Científicas (CSIC) y la Dra. Juliana Sanz Aparicio, Investigador Científico del Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" del Consejo Superior de Investigaciones Científicas (CSIC) certifican que:

Doña Elsa Franco Echevarría, Licenciada en Biología por la Universidad de Salamanca y Máster en Bioquímica, Biología Molecular y Biomedicina por la Universidad de Salamanca, ha realizado bajo nuestra dirección el trabajo titulado "Estudio cristalográfico de proteínas implicadas en la ruta de los inositoles polifosfatados y su relación con el ARN", en el Departamento de Cristalografía y Biología Estructural del Instituto de Química Física "Rocasolano" (CSIC).

En nuestra opinión, este trabajo reúne las condiciones exigidas por la legislación vigente y tiene originalidad, el rigor y la calidad científica necesarios y suficientes para ser presentado por Doña Elsa Franco Echevarría para optar al Grado de Doctor por la Universidad Complutense de Madrid.

Y para que así conste y con los efectos oportunos, firmamos el presente certificado.

Madrid, ____ de _____ de 201__

Fdo: Dra. Beatriz González Pérez
Científico Titular

Fdo: Dra. Juliana Sanz Aparicio
Investigador científico CSIC

A mis padres y hermanas,

AGRADECIMIENTOS

A lo largo de estos años mucha gente ha colaborado en esta Tesis logrando que saliera adelante, a todos aquellos que han hecho posible este trabajo, gracias.

En primer lugar, me gustaría agradecer de manera especial y sincera a mi directora de Tesis, la Dra. Beatriz González toda la confianza depositada en mí. Bea gracias por darme la oportunidad de realizar una Tesis Doctoral en el Roca a tu lado. A lo largo de estos años hemos compartido muchos momentos juntos, gracias por estar siempre disponible, por tu paciencia, por resaltar las cosas buenas, por tu ayuda constante, por tu dedicación, por resolver todas mis dudas, por enseñarme tan bien y por haber sido no solo profesora, directora, investigadora y guía, sino además amiga. Hemos formado un buen equipo juntos y me siento muy afortunada de haberte tenido como directora de Tesis y haber podido estar todos estos años a tu lado. También quiero agradecer especialmente a mi codirectora de Tesis, la Dra. Juliana Sanz Aparicio, gracias por toda tu ayuda, tus ideas, consejos y la gran labor realiza en todo el proceso de esta Tesis.

Quisiera agradecer enormemente a los investigadores principales del Departamento de Cristalografía y Biología Estructural del “Rocasolano” que han ayudado en la realización de este trabajo: Armando, Josemi, Juan, Julia, Lourdes, Martín y MJ. Quisiera reconocer especialmente la labor del Dr. Martín Martínez Ripoll quien siempre está dispuesto a ayudar, con sus clases magistrales y admirables enseñanzas de Cristalografía. Al Dr. Armando Albert de la Cruz y la Dra. María José Sánchez Barrena gracias por vuestra dedicación a que el laboratorio funcione. Armando, gracias por tu brillante idea para la resolución estructural de la IP₅ 2-K. A la Dra. Lourdes Infantes gracias por tu ayuda con los equipos de difracción y por el análisis en la búsqueda de inhibidores. Además, me gustaría agradecer al resto de personas que forman parte del departamento, a Rafa y en especial a Juana, gracias por tu gran labor e infinita ayuda en la Biología Molecular, por resolver mis mil dudas y querer ayudar siempre.

Esta Tesis no hubiese sido posible sin la participación de personas e instituciones que han facilitado y hecho posible este trabajo. Colaboraciones con grupos de alta calidad científica han hecho realidad esta Tesis. Quisiera agradecer al Dr. José Manuel Pérez Cañadillas por ser el motor de Nrd1, por tus enseñanzas y por confiar en mí para este proyecto. Gracias a las Dra. Silvia Zorrilla y la Dra. Begoña Monterroso del CIB por vuestra contribución y ayuda en las medidas de fluorescencia. Gracias al Dr. Charles Brearley por su colaboración con las medidas enzimáticas y permitirme presentar mis avances a su grupo de investigación.

Por supuesto, no me puedo olvidar de los que me empujaron a la aventura de la tesis doctoral y me abrieron las puertas de sus laboratorios. Gracias a la Dra. Henar Valdivieso del IBFG de Salamanca que hizo que me picara el gusanillo de la investigación en el ecuador de la carrera, gracias también a la Dra. Margarita Díaz y al Dr. Ramón Santamaría del IBFG por abrirme en su día las puertas de su laboratorio y permitirme realizar allí la Tesina y el TFM. Gracias a todos vosotros, he podido realizar un Tesis Doctoral. Laura y Sergio, vosotros fuisteis los primeros en enseñarme cómo se trabaja en un laboratorio y os estaré siempre agradecida.

Sin enterarte el tiempo llega y de repente te encuentras preparando la escritura y defensa de la tesis. Han pasado más de cuatro años desde que llegué al Roca y haciendo balance, considero que estos años han sido una época muy especial, llena de metas cumplidas y barreras superadas, sin duda una de las etapas

más felices de mi vida. Aunque a veces ha habido momentos duros, llenos de estrés, de mucho trabajo, de cosas que no salen... el esfuerzo ha merecido la pena y al final llega la recompensa. No me puedo olvidar de todos los compañeros que me han apoyado y ayudado en estos años, todo esto no lo podría haber logrado sin todos vosotros. Gracias por haber formado parte de esta Tesis, espero no olvidarme de nadie.

En el labo pasamos muchas horas juntos y al final nos convertimos en una gran familia, gracias a todos por cada uno de los buenos momentos vividos dentro y fuera del Roca. Quiero agradecer especialmente a todos aquellos compañeros de trabajo que estaban en el Roca en mis comienzos: Ainhoa, Antonio, Ceci, Jorge, M^a Ángela y Sergio. Rocío, gracias por tu comprensión y apoyo en esta etapa final, eres un encanto. A lo que un día fue el 107, no tengo palabras suficientes para agradeceros todo lo que me habéis ayudado y enseñado en estos años. ¡Tere qué buenos momentos pasamos juntas! Noe mil gracias por tu buena disposición constante, por tu ayuda, por tu amistad, por todos los momentos vividos y los viajes realizados. Mer ¡mi oráculo! Hemos pasado más de cuatro años juntas y en esta última etapa tanto tu apoyo como tu ayuda han sido fundamentales, gracias por ser mi confidente, por todos los buenos momentos vividos juntas en los sincros, viajes y cursos. Espero que la amistad que tengo con vosotras perdure en el tiempo y nos veamos de nuevo en futuros viajes y congresos. A mis chicos del 119, Juanlu, eres random y eso te hace especial, hemos compartido terrazas, cines, escapadas ¡gracias! Mi Mayte, qué bien nos hemos entendido, gracias por tu comprensión, tu apoyo sincero, tu ayuda, tu amistad durante estos años y tu risa contagiosa. Siseth y Martín gracias por animar cafés y comidas con vuestro sentido del humor. Al nuevo 107, Isa y Elena, gracias por amenizar el despacho y soportarme en estos últimos meses tan agobiantes. Mucha gente ha pasado por el Roca durante estos años dejando huella, por ello no me puedo olvidar de María, desde que te fuiste, en el Roca ha quedado un huequito vacío porque tienes una alegría especial que contagia a todo el mundo. Hemos compartido muchas cosas dentro y fuera del Roca, como multitud de viajes al sincro que no hubieran sido lo mismo sin ti. Gracias Iván por tu diversión en las quedadas y los buenos momentos vividos. También a todos aquellos que compartís conmigo el día a día en las comidas y los cafés del 115 que siempre recordaré: Patri, Ana, Loreto, Erney, Conchi, Pablo, Ivanna... ¡Gracias!

Por último, a mi familia. En primer lugar, a mis padres, gracias por haberme proporcionado la mejor educación que estaba en vuestras manos y por darme la oportunidad de estudiar lo que realmente me gustaba, gracias a ello descubrí que me apasiona lo que hago. Gracias por enseñarme que con esfuerzo, trabajo y constancia todo se consigue. Gracias por confiar en mis decisiones, aunque sé que resulta difícil comprender qué hago exactamente siempre me habéis apoyado y animado a continuar formándome. A mis hermanas, un pilar fundamental durante todos estos años, las mejores hermanas del mundo, a quienes acudo para solucionar problemas y pedir ayuda. Junto con Foky somos el mejor equipo del mundo. Mi hermana María, aunque nos separen 4500 km de distancia siempre has estado cerca cuando te he necesitado ¡ahora seré la segunda doctora de la familia! A mi otro yo, mi mitad, mi hermana Cristina, gracias por estar ahí en todo momento, por escucharme y ayudarme en el día a día.

Miguel, ya sabes todo lo que significas para mí y lo importante que eres en mi vida. Gracias por entenderme tan bien y ser un apoyo clave durante estos años, por escucharme en todo momento y animarme, por sacar lo mejor de mí día a día. La distancia en los últimos años ha sido dura pero hemos hecho un gran esfuerzo y lo hemos superado con creces! Madrid-Bilbao están a un paso ¡jeeepa! Aunque la cristalografía mola más que la *cryo* me tendré que pasar a tu bando y me tendrás que enseñar. Estoy deseando comenzar una nueva etapa contigo y ver qué nos depara el futuro.

ÍNDICE

RESUMEN	I
SUMMARY	III
ABREVIATURAS	V
1. INTRODUCCIÓN	1
1.1 INOSITOLES POLIFOSFATADOS	3
1.2 PRECURSORES DE LA RUTA DE LOS IPS: PI3K Y RAS	3
1.3 SÍNTESIS DE IPS MEDIANTE LAS IP QUINASAS	5
1.3.1 Familia I: Inositol polifosfato quinasas (IPKs)	6
1.3.1.1 Inositol trisfosfato 3-quinasas (IP ₃ 3-Ks)	7
1.3.1.1.1 Reacción catalítica: sustrato y producto	7
1.3.1.1.2 Isoformas, organización y estructura del dominio catalítico	8
1.3.1.1.3 Implicaciones biológicas	10
1.3.1.2 Inositol polifosfato multiquinasa (IPmKs)	11
1.3.1.2.1 Reacción catalítica	11
1.3.1.2.2 Estructura de IPmK	12
1.3.1.2.3 Funciones de la IPmK	12
1.3.1.3 Inositol pentaquisfosfato 2-quinasa (IP ₅ 2-Ks)	13
1.3.1.3.1 Reacción catalítica e importancia general	13
1.3.1.3.2 Localización y funciones	14
1.3.1.3.3 Estructura de IP ₅ 2-K de planta, reconocimiento de sustrato y mecanismo	14
1.3.1.4 Inositol hexaquisfosfato quinasas (IP ₆ Ks)	17
1.3.1.4.1 Reacción catalítica e IPPs	17
1.3.1.4.2 Isoformas de IP ₆ K y funciones	18
1.3.1.4.3 Interacciones proteína-proteína	19
1.3.1.4.4 Estructura de IP ₆ K en protozoos	19
1.4 INTERFERENCIAS ENTRE LAS RUTAS DE IPS Y BIOGÉNESIS DEL ARN	20
1.4.1 Ruta de IPs y regulación de la transcripción	20
1.4.2 IP ₆ y edición del ARN	21
1.4.3 IP ₅ 2-K en la vigilancia del ARN	21
1.4.4 IP ₆ K e IPmK en exportación del ARN	21
1.5 RECONOCIMIENTO PROTEÍNA-ARN: NRD1 Y PAPEL EN TRANSCRIPCIÓN	22
1.5.1 Motivos de reconocimiento de ARN: dominios “RRM”	22
1.5.2 Nrd1 y terminación de la transcripción	23
1.5.3 Otros componentes del complejo NNS	25
2. OBJETIVOS	27
3. TRABAJOS DE INVESTIGACIÓN	31
3.1 TRABAJOS DE INVESTIGACIÓN 1 Y 2	33
3.1.1 Trabajo de investigación 1: Crystallization and preliminary X-ray diffraction analysis of a mammal inositol 1,3,4,5,6-pentakisphosphate 2-kinase	33
3.1.2 Trabajo de investigación 2: The crystal structure of mammalian inositol 1,3,4,5,6-pentakisphosphate 2-kinase reveals a new zin-binding site and key features for protein function	35
3.1.3 Discusión trabajos de investigación 1 y 2	52

3.2 TRABAJO DE INVESTIGACIÓN 3: THE STRUCTURE OF TRANSCRIPTION TERMINATION FACTOR NRD1 REVEALS AN ORIGINAL MODE FOR GUAA RECOGNITION	57
3.2.1 Discusión trabajo de investigación 3	89
4. MATERIAL INÉDITO	93
4.1 ESTUDIO ESTRUCTURAL DE LA IP ₆ K	95
4.1.1 Introducción	95
4.1.2 Objetivos	95
4.1.3 Materiales y métodos	95
4.1.3.1 Clonación y diseño de construcciones	95
4.1.3.2 Expresión y purificación de IP ₆ K	97
4.1.3.2.1 Expresión y purificación de IP ₆ K recombinante en bacteria	97
4.1.3.2.2 Expresión y purificación de IP ₆ K recombinante en células de insecto	97
4.1.3.3 Obtención de nanoanticuerpos específicos de IP ₆ K2	98
4.1.3.3.1 Generación de nanoanticuerpos frente a IP ₆ K2	98
4.1.3.3.2 Expresión y purificación de los nanoanticuerpos	98
4.1.3.4 Ensayos de cristalización de IP ₆ K, recogida y análisis de datos	99
4.1.3.5 Identificación de las especies cristalizadas	99
4.1.3.6 Modelado de IP ₆ K2	99
4.1.4 Resultados y discusión	100
4.1.4.1 Producción de IP ₆ K2 en <i>E. coli</i> y cristalización	100
4.1.4.2 Producción de IP ₆ K1 en células de insecto y cristalización	102
4.1.4.3 Producción de nanoanticuerpos y formación de complejos con IP ₆ K2	105
4.1.4.4 Modelo	107
4.1.4.4.1 Alineamiento estructural de IP ₆ K, IP ₃ 3-K e IPmK	107
4.1.4.4.2 Modelado de IP ₆ K2 y análisis estructural	108
4.1.5 Conclusiones	110
4.1.6 Perspectivas	110
5. DISUSIÓN GENERAL	113
6. CONCLUSIONES	119
7. BIBLIOGRAFÍA	123
8. ANEXO: MATERIALES Y MÉTODOS	137
8.1 INTRODUCCIÓN A LA BIOLOGÍA ESTRUCTURAL	139
8.2 OBTENCIÓN DE LA PROTEÍNA	140
8.2.1 Clonación y expresión	140
8.2.1.1 Expresión en células de insecto	140
8.2.2 Purificación	142
8.3 CRISTALIZACIÓN DE PROTEÍNAS	143
8.3.1 Concepto de cristal	143
8.3.2 Requisitos previos a la cristalización	144
8.3.3 Principios básicos de cristalización	144
8.3.4 Métodos de cristalización	145
8.3.5 Búsqueda inicial de condiciones de cristalización	146
8.3.6 Optimización de las condiciones de cristalización	147
8.3.7 Cristalización de complejos proteína-ligando	148
8.3.8 Empleo de nanoanticuerpos en la cristalización de proteínas	148
8.4 RECOGIDA DE DATOS DE DIFRACCIÓN DE RAYOS X	149
8.4.1 Elementos de sujeción y protección del cristal	150
8.4.2 Fuentes de rayos X	151

8.4.3 Detección de rayos X	152
8.5 PROCESADO DE DATOS DE DIFRACCIÓN DE RAYOS X Y RESOLUCIÓN ESTRUCTURAL	152
8.5.1 Procesado y escalado de datos	153
8.5.2 Problema de la fase y obtención de mapas de densidad electrónica	154
8.5.2.1 Método de reemplazo molecular (MR).....	155
8.5.2.1 Método de dispersión anómala simple/múltiple (SAD/MAD).....	155
8.5.3 Construcción del modelo estructural y refinado	156
8.5.4 Validación del modelo	157

RESUMEN

Los inositoles polifosfatados (IPs) son un amplio y diverso grupo de pequeñas moléculas solubles que desempeñan papeles biológicos en señalización celular y procesos metabólicos. La ruta de los IPs comienza con la síntesis de inositol 1,4,5-trisfosfato (IP_3), un segundo mensajero muy conocido que promueve la liberación de Ca^{2+} de los sistemas de almacenamiento y regula su homeostasis. El precursor de IP_3 es el inositol lipídico PIP_2 , el cual a su vez es sustrato de la enzima PI3K, uno de los principales efectores de la proteína reguladora de señalización celular RAS. Por tanto, RAS/PI3K participan en la regulación de los niveles de IP_3 . A partir del IP_3 , se forman el resto de IPs mediante la acción de diferentes quinasas y fosfatasas. Esta tesis se centra en el estudio de las inositol polifosfato quinasas (IPKs) de mamíferos, que regulan los niveles de IPs mediante la fosforilación del anillo de inositol en sus diferentes posiciones. En la última década, se ha puesto de manifiesto que muchos papeles de los IPs y las IPKs están directamente relacionados con eventos del ARN y con proteínas que lo reconocen. Este trabajo también se ha centrado en el análisis de una proteína implicada en la terminación de la transcripción en levaduras y a su vez relacionada con la señalización por RAS y otras rutas de IPs/IPKs.

En cuanto a la biología estructural de las IPKs, se han estudiado dos familias diferentes de IPKs mediante Cristalografía de rayos X: la familia de inositol 1,3,4,5,6-pentaquisfosfato 2-quinasa (IP_5 2-K) y la familia de inositol 1,2,3,4,5,6-hexaquisfosfato 6-quinasa (IP_6 K).

La familia IP_5 2-K fosforila el inositol 1,3,4,5,6-pentakisfosfato (IP_5) para producir el inositol 1,2,3,4,5,6-hexakisfosfato (IP_6), a partir de ATP. Esta es la única IPK que fosforila la posición axial del anillo de *myo*-inositol (2-OH), representando un punto clave en la biología de los IPs. Su producto, IP_6 , es el IP más abundante en los seres vivos y está involucrado en múltiples eventos biológicos como la reparación del ADN, exportación y edición del ARN y la regulación de la estructura de la cromatina. Además, es precursor de los inositoles pirofosfatados (IPPs). En los mamíferos, la IP_5 2-K es crítica para el desarrollo embrionario y, aparte de su función catalítica, se ha descrito que IP_5 2-K juega un papel en la regulación de la biogénesis del ARNr mediante interacciones proteína-proteína. Hasta este trabajo, la única estructura conocida de una IP_5 2-K era la de *Arabidopsis thaliana*. Después de muchos esfuerzos, se ha conseguido expresar y purificar la IP_5 2-K de *Mus musculus* (*mIP*₅ 2-K) en bacterias y células de insecto. El diseño de una forma troncada ha permitido resolver la estructura de esta enzima. *mIP*₅ 2-K preserva el plegamiento general de las IPKs, sin embargo, presenta características únicas que se conservan en las distintas isoformas de mamíferos. Estas incluyen elementos novedosos en el reconocimiento y catálisis del inositol, un nuevo sitio de unión de zinc, regiones con un plegamiento diferente y un parche básico en su superficie que podría estar correlacionado con la localización en mamíferos y/o otras posibles funciones en las células.

Finalmente, la familia IP_6 K actúa sobre el IP_6 (u otros IPs completamente fosforilados) y los fosforila en la posición 5 a partir del ATP, produciendo algunos isómeros de IPPs. Los IPPs presentan funciones en procesos apoptóticos, tráfico de vesículas, homeostasis metabólica, respuesta de las células al estrés y diabetes. Los mamíferos expresan tres isoformas de IP_6 K. Se dispone de la estructura de la IP_6 K de *Entamoeba histolytica*, pero esta carece de múltiples elementos presentes en otras isoformas. Por lo tanto, la determinación de la estructura de IP_6 K de mamífero es clave para comprender el mecanismo catalítico de esta familia, su regulación y relaciones evolutivas con otras quinasas. En este trabajo se han producido muestras de IP_6 K de humano plegadas correctamente y diseñado varias estrategias para

cristalizar esta enzima. En paralelo, se ha construido un modelo estructural de IP₆K el cual ha permitido la identificación de residuos clave para la actividad enzimática.

Por otra parte, como se ha mencionado, muchas de las funciones de IPKs e IPs se basan en la interacción directa con vías de biogénesis y metabolismo del ARN. Se ha estudiado un sistema opuesto a la señalización de RAS en *Saccharomyces cerevisiae*, que comprende el complejo NNS, responsable de la terminación de los transcritos no codificantes de la ARN Polimerasa II. Este complejo proteico consta de tres proteínas de unión a ARN: Nrd1, Nab3 y Sen1 (con actividad helicasa de ARN). Específicamente, uno de los papeles de Nrd1 es unirse al ARN naciente que contiene la secuencia GUAA/G. Por Cristalografía de rayos X, se ha determinado la estructura del dominio de unión a ARN de Nrd1. La estructura revela que este dominio a su vez se divide en dos, un RRM canónico y otro denominado dominio “*Split*” con arquitectura novedosa. Esta organización no se ha descrito con anterioridad en dominios RRM. Además, también se ha cristalizado Nrd1 en complejo con diferentes ARNs que contienen GUAA, su secuencia diana. Los resultados permiten proponer un modo único de unión a ARN en el que los dos dominios de Nrd1 están involucrados y amplían nuestro conocimiento sobre los dominios RRM, una de las formas más comunes de unión a ARN en seres vivos.

En conclusión, se ha determinado la estructura de proteínas implicadas en eventos celulares esenciales como señalización celular, control metabólico y transcripción del ARN, permitiendo un gran avance en el entendimiento de sus funciones y aportando herramientas valiosas para el diseño de inhibidores.

SUMMARY

The inositol polyphosphates (IPs) are a diverse group of small soluble molecules that play biological roles in metabolic, cellular and signalling processes. The IPs signalling pathway starts with the synthesis of inositol 1,4,5-triphosphate (IP₃), a well known second messenger that promotes Ca²⁺ release from intracellular stores and regulates Ca²⁺ homeostasis. The IP₃ precursor is PIP₂, a lipid inositol substrate of PI3K, which is a main RAS effector that orchestrates cell signalling. Therefore, the complex RAS/PI3K participates in IP₃ levels regulation. From IP₃, other IPs are formed successively by the action of different kinases and phosphatases. The main goal of this work is focused in the study of the mammal inositol polyphosphate kinases or IPKs that regulate IPs levels by phosphorylating the inositol ring at its different positions. In the last decade, it has become evident that main roles of IPs and IPKs are directly related with RNA events. In particular, this work has also been focused on a protein involved in transcription termination in yeasts that is related to RAS signalling and presents cross-talks with IPs/IPKs-participating routes.

Respect to the IPKs, two different IPKs have been studied by X-ray Crystallography: inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IP₅ 2-K) and inositol 1,2,3,4,5,6-hexakisphosphate-kinase (IP₆K).

The IP₅ 2-K family phosphorylates inositol 1,3,4,5,6-pentakisphosphate (IP₅) to produce inositol 1,2,3,4,5,6-hexakisphosphate (IP₆), from ATP. IP₅ 2-K is the unique IPK that phosphorylates the 2-OH axial position of the *myo*-inositol ring, representing a hotspot in inositol biology. Its product, IP₆, is the most abundant IP in living beings and is involved in multiple biological events as DNA repair or regulation of chromatin structure. Moreover, it is a cofactor for different RNA-editing enzymes and is the precursor of pyrophosphate inositols (IPPs). In mammals, IP₅ 2-K is involved in embryonic development and is linked to proteins ubiquitination. Apart from its catalytic function, a structural role in the cell nucleolus has been described for IP₅ 2-K, interacting with proteins that regulate rRNA biogenesis. Up to this now, the only IP₅ 2-K known structure was that from *Arabidopsis thaliana*. After many efforts, we were able to express and purify IP₅ 2-K from *Mus musculus* (mIP₅ 2-K) in bacteria and insect cells. To get crystals of good quality a truncated protein form was designed, which finally allowed us to solve the structure of mIP₅ 2-K. mIP₅ 2-K structure preserves the general folding of IPKs and shares characteristics with the plant enzymes. Nevertheless, it presents unique features that are conserved in the mammal isoforms. These include unique elements for the inositol recognition and catalysis, a novel zinc binding site, different folding in some regions, and a basic patch on its surface which could be correlated with mammal localization and/or other possible functions in cells.

Finally, the IP₆Ks act mainly on IP₆ phosphorylating its 5-position from ATP, yielding some IPPs isomers. IPPs are involved among other functions in apoptotic processes, cellular trafficking, metabolic homeostasis, response of cells to stress and diabetes. Mammals express three IP₆Ks isoforms. The structure of the *Entamoeba histolytica* IP₆K is available, but it lacks multiple elements present in other identified isoforms. Therefore, the determination of a mammalian IP₆K structure is a major goal in order to understand its catalytic mechanism, regulatory processes and evolutionary relationships with other kinases. We have produced IP₆K samples properly folded and designed several strategies to crystallize this enzyme. In parallel, we have built an IP₆K structural model that allowed the identification of key residues for enzymatic activity.

On the other hand, as mentioned above, many of the IPKs and IPs functions are based on their direct interaction with the RNA biogenesis pathways. Therefore, our interest has been focused on the study of this relationship and RNA binding proteins. To start with, we have chosen a system comprising the NNS complex, which shows opposite effects to RAS signalling in *Saccharomyces cerevisiae*, and controls the termination of the non-coding RNA transcripts from RNA Polymerase II. This protein complex consists of three RNA-binding proteins: Nrd1, Nab3 and Sen1 (with RNA helicase activity). Specifically, one of Nrd1 roles is to bind nascent RNAs, which contain the GUAA/G sequence. By X-ray Crystallography, we have been able to determine the structure of the RNA binding domain of Nrd1. The structure reveals that, in fact, Nrd1 presents two domains: a canonical RRM and a novel domain (Split domain) with a new architecture. This organization has not been described before for the RRM domains. Moreover, we have also crystallized Nrd1 bound to different RNA complexes containing GUAA, its target sequence. Our results allowed us to propose a unique mode of RNA binding, in which the two domains of Nrd1 are involved. This expands our knowledge on RRM domains, one of the most common ways of RNA binding in living things.

In conclusion, the structure of several proteins involved in key cellular events as cell signaling, metabolic control and RNA transcription have been determined. The results provide knowledge on this proteins functions and represent a valuable tool to drug design.

ABREVIATURAS

ABD: Dominio de unión a Actina (“*Actin Binding Domain*”)

ADN: Ácido desoxiribonucleico (“*DNA, Deoxyribonucleic acid*”)

ADNc: ADN complementario (“*DNAc, Complementary DNA*”)

ADP: Adenosin difosfato (“*Adenosine DiPhosphate*”)

Amp: Ampicilina

AMP PNP: Adenosina 5'(β,γ -imido)trifosfato

ARN: Ácido ribonucleico (“*RNA, Ribonucleic acid*”)

ARNm: **ARN** mensajero

ARNnc: **ARN** no codificante

ARN Pol: **ARN** Polimerasa

ARNr: **ARN** ribosómico

ATP: Adenosin trifosfato (“*Adenosine TriPhosphate*”)

BEVS: Sistema de expresión en baculovirus (“*Baculovirus Expression Vector System*”)

BME: 2-mercaptoetanol o β -mercaptoetanol

CaM: Calmodulina (“*Calcium Modulated Protein*”)

CARD: Dominios de reclutamiento de caspasas (“*Caspase Recruitment Domains*”)

CBD: Dominio de unión a calmodulina (“*Calmodulin Binding Domain*”)

CCD: Dispositivo de carga a acoplada (“*Charged-Coupled Device*”)

CID: Dominio de interacción con CTD (“*CTD Interacting Domain*”)

CK2: Caseína Quinasa II (“*Casein Kinase II*”)

CTD: Dominio C-terminal de la Pol II (“*C-terminal Domain*”)

CUTs: Transcritos crípticos inestables (“*Cryptic Unstable Transcripts*”)

Da: **Dalton**

DAG: **Diacilglicerol**

EDTA: Ácido etilendiamino tetra-acético (“*Ethylenediaminetetraacetic acid*”)

ELA: **E**slclerosis **L**ateral **A**miotrófica

DIPP: Difosfo inositol polifosfato difosfatasa (“*Diphosphoinositol-polyphosphate diphosphatase*”)

DTT: **D**itiotreitol

ESRF: Instalación europea de radiación sincrotrón (“*European Synchrotron Radiation Facility*”)

FYVE: **F**ab1, **Y**OTB, **V**ac1 **E**EA1

GST: **G**lutatión-**S**-transferasa

HSP90: Proteína de shock térmico de 90 kDa (“*Heat Shock Protein 90*”)

I(1,4)P₂/IP₂: Inositol 1,4-bisfosfato

I(1,4,5)P₃ o IP₃: Inositol 1,4,5-trisfosfato
 I(1,3,4)P₃: Inositol 1,3,4-trisfosfato
 I(1,3,4, 5)P₄ o IP₄: Inositol 1,3,4,5-tetraquisfosfato
 I(3,4,5,6)P₄: Inositol 3,4,5,6-tetraquisfosfato
 I(1,3,4,5,6)P₅ o IP₅: Inositol 1,3,4,5,6-pentaquisfosfato
 I(1,2,3,4,5,6)P₆ o IP₆: Inositol 1,2,3,4,5,6-hexaquisfosfato
 1-IP₇: 1-bifosfoinositol pentaquisfosfato
 5-IP₇: 5-bifosfoinositol pentaquisfosfato
 1,5-IP₈: 1,5-bifosfoinositol tetraquisfosfato
 IPTG: **I**sopropil- β -D-**t**iogalactósido
 IP(s): Inositol **p**olifosfato(s) (“*Inositol polyphosphate(s)*”)
 IPK(s): Inositol polifosfato quinasa(s) (“*Inositol polyphosphate kinase(s)*”)
 IPP(s): Inositol pirofosfatado(s) (“*Inositol pyrophosphate(s)*”)
 IPMK: Inositol polifosfato **m**ultiquinasa
 IP₃ 3-K: Inositol 1,4,5-trisfosfato 3-quinasa
 IP₃ 5/6-K: Inositol 1,3,4-trisfosfato 5/6-quinasa
 IP₅ 2-K: Inositol 1,3,4,5,6-pentaquisfosfato 2-quinasa
 IP₆K: Inositol 1,2,3,4,5,6-hexaquisfosfato quinasa
 ITC: Calorimetría isotérmica de titulación (“*Isothermal titration calorimetry*”)
 ITPK1: Inositol 1,3,4-trisfosfato 5/6-quinasa (“*Inositol-Tetrakisphosphate 1-Kinase*”)
 kDa: **K**ilodalton
 LSL: Lectina de *Laetiporus sulphureus* (“*Laetiporus sulphureus lectin*”)
 MAD: Difracción anómala múltiple (“*Multi-wavelength anomalous dispersion*”)
 MBP: Proteína de unión a maltosa (“*Maltose Binding Protein*”)
 PCT: Test de pre-cristalización (“*Pre-Crystallization Test*”)
 PDB: Banco de Datos de Proteínas (“*Protein Data Bank*”)
 PEG: **P**olietilenglicol
 PEG 3350: **P**olietilenglicol **3350** Da
 PEG 6K: **P**olietilenglicol **6** kDa
 PI3K: Fosfoinositol 3-quinasa (“*Phosphoinositide 3-kinase*”)
 PIP₂: Fosfatidilinositol 4,5-bisfosfato
 PIP₃: Fosfatidilinositol 3,4,5-trisfosfato
 PK(s): **P**roteína(s) quinasa(s) (“*Protein Kinase(s)*”)
 PKA: Proteína quinasa A (“*Protein Kinase A*”)
 PKC: Proteína quinasa C (“*Protein Kinase C*”)

PLC: Fosfolipasa C (“***P**hospholipase **C***”)

PIP5K: Difosfoinositol pentaquisfosfato quinasa (“***D**iphosphoinositol pentakisphosphate kinase*”)

PMSF: Fluoruro de fenilmetilsulfonilo (“***P**henyl**m**ethylsulfonyl **f**luoride*”)

RMN: **R**esonancia **m**agnética **n**uclear

RNP: Ribonucleoproteína (“***R**ibonucleoprotein*”)

RRM: Motivo de reconocimiento de ARN (“***R**NA **R**ecognition **M**otif*”)

SAD: Dispersión anómala simple (“***S**inlge wavelength **a**nomalous **d**ispersion*”)

Se-Met: **S**elenio **M**etionina

SD: Dominio dividido (“***S**plit **D**omain*”)

SDS-PAGE: Electroforesis desnaturalizante en gel de poliacrilamida (“***S**DS-**P**olyacrylamide **g**el **e**lectrophoresis*”)

SUTs: transcritos estables sin anotar (“***S**tale **u**nannotated **t**ranscripts*”)

TEV: (“***T**obacco **E**tch **V**irus*”)

UBF: “***U**pstream **b**inding **f**actor*”

1. INTRODUCCIÓN

1.Introducción

1.1 Inositoles polifosfatados

El inositol o el ciclohexano-1,2,3,4,5,6-hexol es un carbohidrato que se produce ampliamente en la naturaleza. Posee estructura de anillo formado por seis grupos de carbono, cada uno de los cuales tiene un hidroxilo libre. Existen nueve estereoisómeros de inositol: *scyllo*, *myo*, *epi*, *chiro* (*D* o *L*), *neo*, *allo*, *cis* y *muco*, de los cuales el *myo*-inositol es la forma más abundante (Figura 1.1). El *myo*-inositol, en la conformación de silla, presenta el grupo hidroxilo en la posición 2 axial, esto es perpendicular al plano del anillo, mientras que los otros cinco grupos hidroxilo están en posición ecuatorial, es decir aproximadamente en el mismo plano que el anillo. Como consecuencia de la stereoquímica descrita, el *myo*-inositol posee un eje de simetría a través de sus posiciones 2 y 5. El *myo*-inositol forma la base de la familia de los llamados inositoles polifosfatados (IPs) los cuales se sintetizan sustituyendo los grupos hidroxilo por grupos fosfato. Existen 63 variedades posibles de IPs y hasta la fecha se han identificado más de 30 en las células eucariotas. Los IPs son moléculas pequeñas y solubles que juegan un papel fundamental en procesos celulares, metabólicos y de señalización como la regulación de canales iónicos, transcripción, desarrollo embrionario, reparación de ADN o apoptosis entre otras funciones (Saiardi, 2012; Shears et al., 2012).

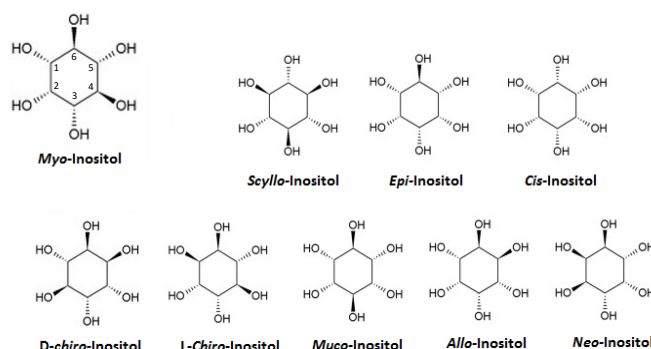


Figura 1.1. Isómeros de inositol. La figura representa los 9 isómeros de inositol existentes. Se destaca el anillo de *myo*-inositol con los carbonos numerados.

Los niveles de los IPs se encuentran regulados por diferentes familias de IP quinasas o IP fosfatasas que fosforilan/desfosforilan respectivamente las diferentes posiciones del anillo de inositol.

En las sucesivas secciones se tratarán en detalle las diferentes familias y subfamilias que forman parte del complejo metabolismo de los IPs, así como el inicio de las rutas que llevan a su síntesis.

1.2 Precursores de la ruta de los IPs: PI3K y RAS

El punto de partida para la síntesis de los IPs se encuentra ligado a proteínas cruciales en señalización celular y cáncer: fosfoinositol 3-quinasa (PI3K) y RAS (Figura 1.2).

La enzima PI3K, convierte el fosfolípido fosfatidilinositol 4,5-bisfosfato (PIP₂) en el segundo mensajero fosfatidilinositol 3,4,5-trisfosfato (PIP₃), controlando el balance entre ambos fosfolípidos de inositol (Walker et al., 1999). PIP₂ posee un papel fundamental en la ruta de los IPs, ya que también es el precursor del segundo mensajero inositol 1,4,5-trisfosfato (I(1,4,5)P₃ o IP₃). Para ello, el PIP₂ actúa como sustrato para la fosfolipasa C (PLC), enzima que lo escinde en IP₃ y diacilglicerol (DAG) (Quinlan, 2006).

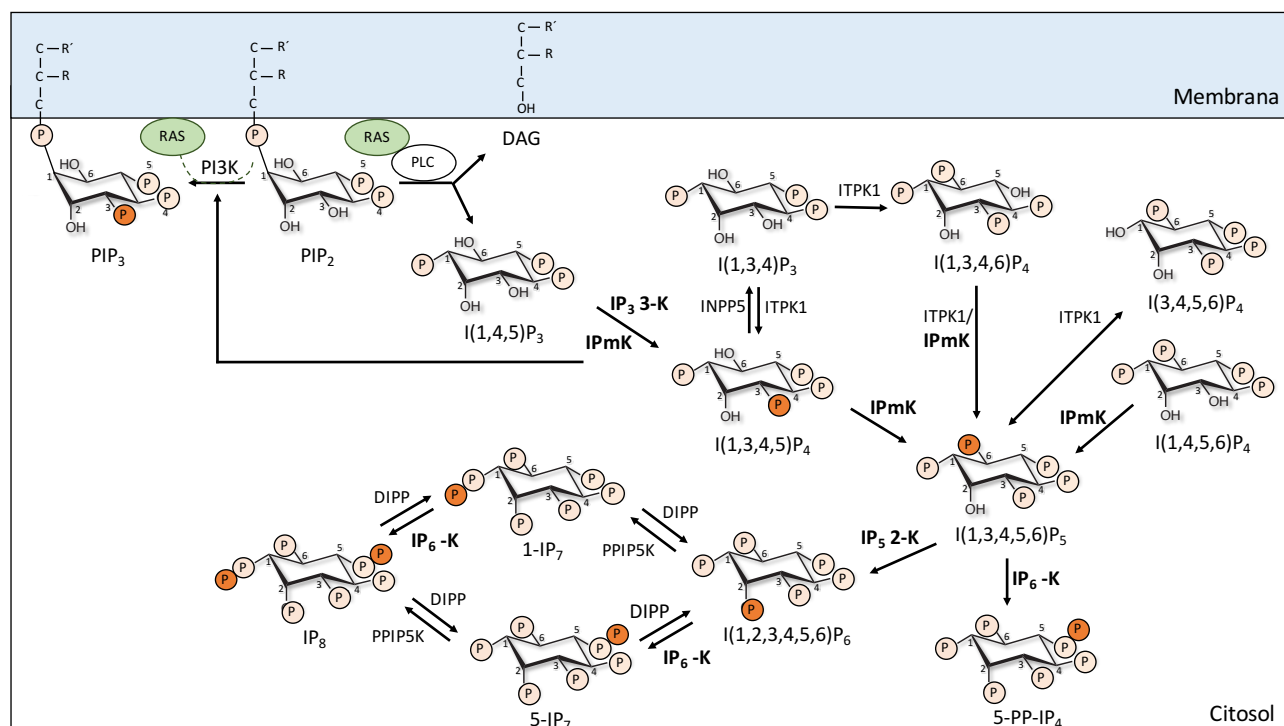


Figura 1.2. Esquema de la ruta de los IPs en mamíferos. En negrita se resaltan las enzimas pertenecientes a la familia IPK, y en naranja los fosfatos fruto de las diferentes reacciones de fosforilación.

Por otro lado, la proteína RAS es un activador de la familia de PI3Ks. RAS es una proteína G monomérica, con actividad reguladora GTP-hidrolasa. Alterna dos conformaciones estructurales: en su forma activa, RAS está unida a GTP mientras que en su forma inactiva se encuentra unida al GDP. PI3K es uno de los efectores más importantes de RAS, y ambas proteínas regulan una gran variedad de rutas de señalización celular, controlando diferentes procesos como la proliferación y supervivencia celular, diferenciación, adhesión, apoptosis y transcripción (Pacold et al., 2000). En cáncer, mutaciones de PI3K y RAS desencadenan un estado constitutivamente activado de las enzimas, que se traduce en un aumento de la capacidad de invasión y metástasis, y una disminución de la apoptosis. Por este motivo, tanto RAS como PI3K son dianas importantes en la lucha contra el cáncer (Papke y Der, 2017).

Por tanto, PIP₂, el precursor de IP₃ y en consecuencia el punto de partida de la ruta de los IPs, juega un papel central en señalización celular al ser el precursor de diferentes segundos mensajeros. A partir de IP₃ se producen sucesivas reacciones de fosforilación/defosforilación catalizadas por las IP quinasas y fosfatasas (Figura 1.2).

1.3 Síntesis de IPs mediante las IP quinasas

Las IP quinasas son las enzimas encargadas de la síntesis de los diferentes IPs en combinación con las IP fosfatasas. Existen dos familias estructurales de IP quinasas, las cuales presentan un dominio catalítico con actividad quinasa en algunos casos unido a otros dominios con funciones principalmente reguladoras y/o de localización. Estas dos familias son:

- Familia I: Inositol polifosfato quinasas (IPKs).
- Familia II: Quinasas con plegamiento “ATP-*grasp*”.

Estas dos familias, IPKs y “ATP-*grasp*” cubren todas las reacciones de fosforilación en las diferentes posiciones del anillo de inositol. Este trabajo se ha centrado en el estudio de proteínas pertenecientes a la familia I, la cual se describe en detalle más adelante.

Respecto a la familia II, únicamente mencionar que existen dos representantes dentro de esta familia: inositol 1,3,4-trisfosfato 5/6-quinasa (IP₃ 5/6-Ks o ITPK1) y difosfoinositol pentaquisfosfato quinasa (PIP5K) (Figura 1.2). La familia ITPK1 posee tanto actividad 1/5/6-quinasa como fosfatasa. Se encuentra conservada desde las plantas a los seres humanos y regula múltiples funciones celulares como la secreción de fluidos e insulina (Ho et al., 2002). En cuanto a la familia PIP5K, es responsable de la síntesis de los inositoles pirofosfatados (IPPs): 1-IP₇ (1-bisfosfoinositol pentaquisfosfato) e 1,5-IP₈ (1,5-bisfosfoinositol tetraquisfosfato) (Wang et al., 2014b). Esta enzima posee actividad 1-quinasa comenzando a actuar sobre el IP₆ (Draskovic et al., 2008) y se encuentra conservada de levaduras a humanos.

En 2005 se determinó la primera estructura de la familia II, ITPK1 del organismo *Entamoeba histolytica* (código PDB 1Z2N) (Miller et al., 2005) (Figura 1.3) y posteriormente en 2007 se resolvió la estructura de la isoforma de *Homo sapiens* (código PDB 2Q7D) (Chamberlain et al., 2007). Más recientemente, en 2011, se conoció la estructura del dominio catalítico de PIP5K2 (código PDB 3T9C) (Wang et al., 2011). En general, esta familia de quinasas presenta un plegamiento y reconocimiento de ligando diferente a las IPKs.

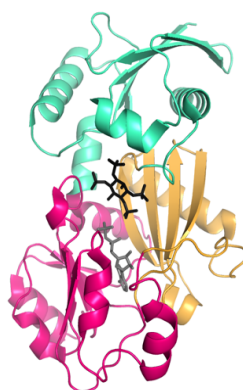


Figura 1.3. Estructura de ITPK1 de *Entamoeba histolytica* (código PDB: 1Z2O). Estructura representada en diagrama de cintas: dominio N-terminal en verde, dominio C-terminal en amarillo y el dominio central en rosa. El IP₄ se representa en negro y el ATP en gris. ITPK1 fue la primera estructura reportada de la familia II de IPKs.

1.3.1 Familia I. Inositol polifosfato quinasas (IPKs)

Dentro de la familia I o IPKs existen cuatro subfamilias que se clasifican en función de las reacciones que catalizan (Figura 1.2):

- Inositol trifosfato 3-quinasas (IP_3 3-Ks)
- Inositol polifosfato multiquinasas (IPmKs)
- Inositol pentaquisfosfato 2-quinasas (IP_5 2-Ks)
- Inositol hexaquisfosfato quinasas (IP_6 Ks)

A pesar de su baja conservación de secuencia, todas las IPKs comparten un plegamiento general similar organizado en dos lóbulos N- y C-terminal, conectados por un bucle de longitud y secuencia variable. Además, todas reconocen entre los dos lóbulos y de manera similar el ATP, el donador de fosfato, que a su vez está relacionado con el reconocimiento en las proteínas quinasas (PKs) (Figura 1.4) (Cheek et al., 2005).

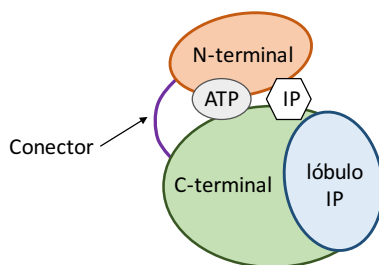


Figura 1.4. Esquema del plegamiento de una IPK. En naranja se representa el lóbulo N-terminal, en verde el C-terminal y en azul el lóbulo de unión a inositol (lóbulo IP). En morado el bucle conector entre lóbulos N- y C- terminal. Además, se muestran en gris y blanco, las regiones de unión al nucleótido y al IP respectivamente.

En particular, el lóbulo N-terminal conserva el plegamiento básico α y β de las proteínas quinasas (PKs), superfamilia a la cual pertenecen las IPKs, si bien algunas IPKs presentan inserciones en esta región. Este plegamiento consta de una lámina β formado por 5 hojas β antiparalelas con orden 1-2-3-5-4, y una hélice insertada entre las β_3 y β_5 denominada hélice α_C (Figura 1.5), esencial esta última en las PKs por ser responsable de la formación de su conformación activa. Entre las hojas β_1 y 2 está el bucle G, esencial para la coordinación de los fosfatos del ATP (Figura 1.5).

Respecto al lóbulo C, este posee un plegamiento específico y nuevo en las IPKs, presentando, al contrario que en las PKs, un núcleo de hojas β con inserciones de diferentes longitudes dependiendo de la clase de IPK. Este plegamiento junto con los elementos implicados en el mecanismo catalítico de las IPKs se describirá en detalle más adelante.

En cuanto al sitio de unión al inositol, este se encuentra localizado en una región formada por hélices α insertada en el lóbulo C (lóbulo de unión a IP) (Figura 1.4). Las cuatro familias de IPKs presentan sus máximas diferencias en esta región de unión al inositol, siendo característica para cada familia. En la IP_3 3-K, el dominio de unión a inositol consiste en una región de 60 residuos organizados en cuatro hélices α , mientras que, tanto en las IP_6 Ks como en las IPmKs, el dominio de unión a inositol es mucho más corto, abarcando aproximadamente 30 residuos que forman una única hélice. En cuanto a la IP_5 2-K, esta presenta la región de unión a inositol más grande dentro de las IPKs abarcando casi la mitad del

tamaño de la enzima (más de 150 residuos formando once hélices α). Ello se traduce en que estas poseen distinta especificidad por el sustrato. Mientras que las IP₃ 3-Ks y las IP₅ 2-Ks son muy específicas en cuanto al sustrato que unen, las IPmKs e IP₆Ks muestra actividad catalítica frente a más de un sustrato. Por otro lado, mientras que las IP₃ 3-Ks como las IP₅ 2-Ks y las IP₆Ks fosforilan una determinada posición del anillo, las IPmKs lo hacen en más de una posición (Bennett et al., 2006; Gonzalez et al., 2004; Miller y Hurley, 2004; Wang et al., 2014a).

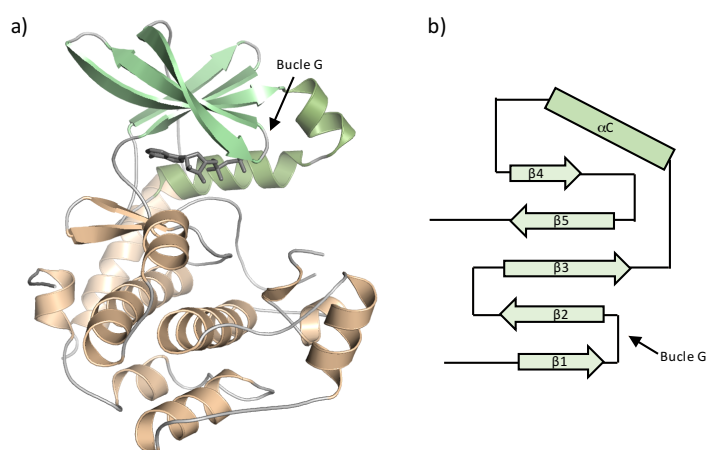


Figura 1.5. Estructura de la proteína quinasa A (PKA) (código PDB 4WB5). a) Estructura general de PKA, como representante de la familia de PKs, representada en diagrama de cintas con el código de colores: lóbulo N-terminal en verde, lóbulo C-terminal en crema. Se indica el bucle-G conservado en las IP quinastas. b) Diagrama de topología del lóbulo N. Se representan las 5 hojas β antiparalelas y la hélice αC ; con una flecha se indica el bucle-G.

A día de hoy, se conoce al menos la estructura de un miembro de cada subfamilia de IPKs. En 2004 se resolvió la primera estructura de una IPK, correspondiente a la isoforma A de la IP₃ 3-K de humano (código PDB 1W2C) (Gonzalez et al., 2004). Posteriormente se obtuvo la estructura de IPmK de levadura (código PDB 2IEW) (Holmes y Jogl, 2006). Más recientemente, en 2010, nuestro grupo de investigación resolvió la estructura de la IP₅ 2-K de *Arabidopsis thaliana* (código PDB 2XAN) (Gonzalez et al., 2010), lo que demostró que esta enzima pertenece a la familia IPK aunque es el miembro más distante. Finalmente, en 2014 se ha conocido la estructura de una IP₆K perteneciente a *Entamoeba histolytica* (código PDB 4O4E) (Wang et al., 2014a).

1.3.1.1 Inositol trifosfato 3-quinastas (IP₃ 3-Ks)

1.3.1.1.1 Reacción catalítica: sustrato y producto

La IP₃ 3-K cataliza la fosforilación del inositol 1,4,5-trisfosfato, I(1,4,5)P₃ o IP₃ en la posición 3 del anillo, formando el inositol 1,3,4,5-tetraquisfosfato, I(1,3,4,5)P₄ o IP₄ (Figura 1.6).

El sustrato, IP₃, es un segundo mensajero de la transducción de señales en células animales. Su principal función es la liberación del Ca²⁺ del retículo endoplasmático al citosol. Para ello se une a receptores específicos de IP₃ localizados en la membrana externa del retículo endoplasmático que permiten la liberación de Ca²⁺ (Berridge y Irvine, 1984; Seo et al., 2012). Además, el IP₃ sirve como precursor para la síntesis de los restantes IPs (Figura 1.2). El IP₃ también puede ser metabolizado por la inositol polifosfato 5-fosfatasa (INPP5) dando lugar al inositol 1,4-bisfosfato, I(1,4)P₂ o IP₂.

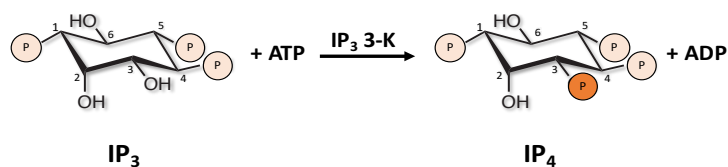


Figura 1.6. Reacción catalizada por la IP₃ 3-K. Síntesis de IP₄ a partir de IP₃ y ATP.

Por su parte el producto de la reacción, el IP₄, es otro mensajero responsable de mediar la entrada de Ca²⁺ a través de la membrana plasmática y movilizar el Ca²⁺ intracelular actuando sinérgicamente con el IP₃, ya que juntos regulan la homeostasis del Ca²⁺. Este también posee un papel como un mediador en la muerte neuronal en el hipocampo isquémico (Tsubokawa et al., 1994) y está involucrado en el desarrollo celular inmune. En particular, la eliminación de la isoforma B en ratones produce fallos en el desarrollo, proliferación y producción de citoquinas en las células T, precisamente por la falta de síntesis del IP₄ (Pouillon et al., 2013).

1.3.1.1.2 Isoformas, organización y estructura del dominio catalítico

Hasta la fecha, esta enzima se ha identificado en numerosos metazoos (Irvine y Schell, 2001; Pattni y Banting, 2004). Existen tres isoformas distintas de IP₃ 3-K codificadas por el genoma humano, las cuales presentan diferentes masas moleculares (51, 100 y 75 kDa respectivamente), así como desigual localización y expresión tisular. Mientras que la A y la B son exclusivamente citosólicas, la C se encuentra también en el núcleo. La isoforma A tiene un claro predominio en neuronas y testículos, por el contrario, las isoformas B y C tienen una expresión ubicua (Chamberlain et al., 2005; Dewaste et al., 2003). Respecto a su organización estructural, IP₃ 3-K se divide en dos regiones, catalítica y reguladora.

La región catalítica constituye un dominio en sí mismo con actividad quinasa situado en el extremo C-terminal (Figura 1.7 y 1.9) y se encuentra muy conservado en las tres isoformas: A (código PDB 1W2C), B (código PDB 2AQX) y C (código PDB 2A98). Este dominio presenta el plegamiento general de las IPKs. De hecho, el dominio catalítico de la isoforma A fue la primera IPK cuya estructura se dio a conocer y por tanto la cabeza de serie a la hora de describir el plegamiento de esta familia. Por un lado, el lóbulo N-terminal conserva el plegamiento de las IP quinastas como se ha mencionado anteriormente (Figura 1.5). Este consta de un núcleo de hojas β antiparalelas con orden 2-3-5-4 junto con una hélice α (equivalente a la hélice α C en PKs). Se observa que carece de la primera hoja β presente en las PKs (Figura 1.5 y 1.7). Por otro lado, el C-terminal está formado por un núcleo de cinco láminas β antiparalelas, en orden 5-4-1-2-3 entre las cuales se insertan diferentes regiones helicoidales. La mayor de estas inserciones está formada por cuatro hélices que construyen el lóbulo de unión a inositol (Figura 1.7).

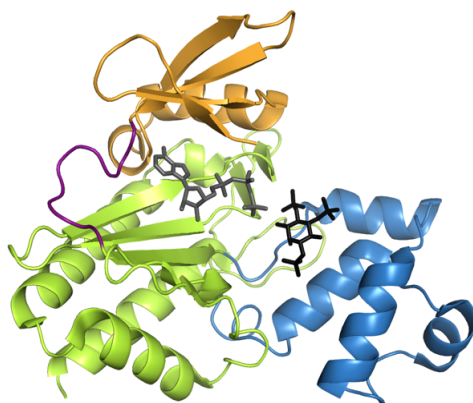


Figura 1.7. Estructura del dominio catalítico de la IP₃ 3-K A de *Homo Sapiens* (código PDB: 2WC). Estructura general de IP₃ 3-K representada en diagrama de cintas siguiendo el siguiente código de colores: lóbulo N-terminal en naranja, lóbulo C-terminal en verde y el lóbulo IP en azul. El conector entre los lóbulos N y C-terminal aparece en violeta. El IP₃ se representa en negro y el ATP en gris.

La obtención de complejos cristalinos de la IP₃ 3-K en presencia de sustratos y productos, reveló todos los detalles del reconocimiento de ATP y de IP. Como se ha mencionado, el ATP se une entre ambos lóbulos (N- y C-) y es reconocido mediante residuos polares e hidrofóbicos (Figura 1.8).

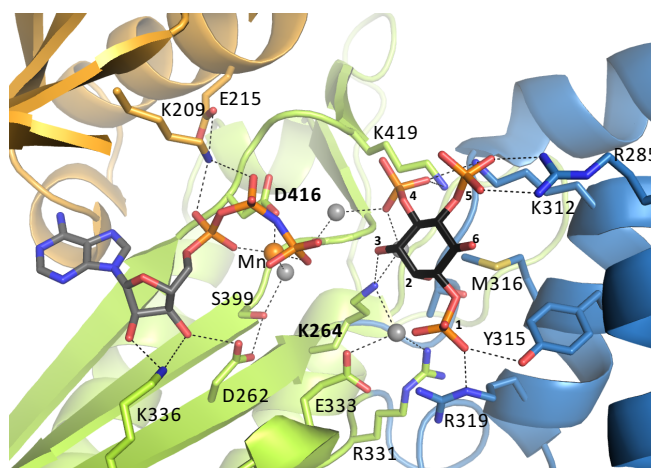


Figura 1.8. Centro activo de la IP₃ 3-K. Detalle de los residuos implicados en el reconocimiento del análogo de ATP (AMP PNP) e IP₃. En negrita se resaltan los residuos catalíticos Asp416 y Lys264. Las moléculas de agua se representan en esferas grises y el ión Mn²⁺ en naranja.

En la unión de los fosfatos del ATP juega un papel clave un bucle flexible del lóbulo N-, denominado bucle G por su analogía con las PKs y su alto contenido en glicinas. Por otro lado, el IP₃ se encuentra muy unido a la enzima en su mayoría mediante residuos básicos situados en el lóbulo de unión a IP, si bien también participa algún residuo fuera del mismo. Esta unión tan fuerte explica el valor de la Km para el IP₃ (1.1 μM) y porqué esta enzima es tan específica. Así, el centro activo únicamente permite la unión del IP₃ y en una posición determinada, exponiendo al ATP únicamente el 3-OH. En cuanto al mecanismo catalítico, se propone que la reacción de transferencia del fosfato gamma (P_γ) de ATP al IP₃ es directa, mediante un mecanismo de sustitución nucleófila bimolecular (SN₂) como ya se ha determinado en las PKs. La proximidad de P_γ (ATP) y 3-OH (IP₃) en las estructuras cristalinas es

compatible con el mecanismo de transferencia directa. La estructura permitió identificar un par de residuos clave en la catálisis: un par ácido/básico situado en el lóbulo-C con papeles muy definidos (Figura 1.8). El residuo ácido (Asp416) orienta el Py a través de un ión Mg^{2+} , mientras que el residuo básico (Lys264) neutraliza la carga negativa que se desarrolla durante la catálisis. Estos residuos están conservados en todas las IPKs y PKs.

En cuanto a la región reguladora, esta se encuentra menos conservada y se ha identificado la presencia de dos dominios (Figura 1.9). Por un lado, el extremo N-terminal es altamente divergente y presenta un dominio de unión a F-actina en las isoformas A y B, que gobierna tanto la localización celular como las funciones específicas de cada isoforma (Communi et al., 1999; Dewaste et al., 2003; Irvine y Schell, 2001). Así, la isoforma A se asocia con la F-actina de las espinas dendríticas donde controla su morfología (Koster et al., 2016; Schell et al., 2001). Igualmente, la isoforma B se asocia con el citoesqueleto a través de F-actina, pero además se asocia con el retículo endoplasmático (Pattni y Banting, 2004). Por otro lado, dentro de la región reguladora, existe una región conservada en la zona central de la enzima, identificado como un dominio de unión a calmodulina ("*Calcium Modulated Protein*" o "CaM") denominado CBD ("*Calcium Binding Domain*"). Las tres isoformas de mamíferos pueden ser activados por CaM de una manera dependiente del ión Ca^{2+} . Esta regulación por Ca^{2+} /CaM es diferente para cada isoforma. Así, en la isoforma A, la unión de CaM induce un aumento de la actividad del dominio quinasa entre 2-3 veces por encima del nivel basal. En cuanto a la isoforma B este aumento es mayor, en torno a 7-8 veces. Por último, la isoforma C es activada por CaM en un grado menor que las isoformas A o B (1.5 veces) (Dewaste et al., 2000). Existen otros mecanismos de regulación de la actividad quinasa de IP_3 3-K además de la regulación por CaM. De este modo, la PKC es capaz de fosforilar las tres isoformas, produciendo una inhibición de la actividad enzimática (Lin et al., 1990). Además, las isoformas A y B son fosforiladas por la CaM quinasa II (CamKII) y la proteína quinasa A (PKA) produciendo un aumento en su V_{max} (Communi et al., 1997) (Figura 1.9).

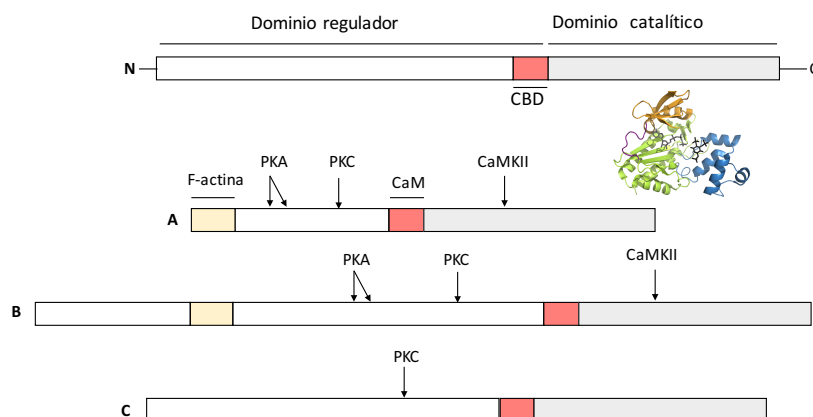


Figura 1.9. Organización general de IP_3 3-K en dominio regulador y catalítico. Esquema general de una IP_3 3K y de las tres isoformas en detalle, donde las regiones amarillas indican las zonas de interacción con F-actina, en coral las zonas reguladoras de interacción con CaM y en gris el dominio catalítico quinasa.

1.3.1.1.3 Implicaciones biológicas

Además de las funciones concretas del sustrato y producto de esta enzima anteriormente detalladas, se ha reportado la intervención de la IP_3 3-K en multitud de procesos biológicos. Así, esta enzima participa en la exportación de ARNm del núcleo en levaduras (Odom et al., 2000); protege frente a daños

oxidativos en respuesta a estrés en *Drosophila* (Monnier et al., 2002); está asociada con la germinación del polen y el crecimiento de las raíces en planta (Xu et al., 2005) y se la relaciona con procesos de desarrollo y diferenciación muscular en mamíferos (Carrasco et al., 1997), así como con la respuesta inmune y la señalización neuronal (Schell, 2010). En particular, la isoforma A, desempeña un papel fundamental en el procesamiento y expresión de los estados afectivos en mamíferos y se le atribuye una posible función en procesos de memoria (Mailleux et al., 1991). Por otro lado, la expresión temprana de esta isoforma en tumores incrementa el potencial metastásico de células de cáncer de pulmón (Windhorst et al., 2011). De hecho, su relación con la carcinogénesis, sugiere que esta enzima es una buena candidata como marcador molecular en la medicina clínica (Kato et al., 2006). En cuanto a la isoforma B, esta controla la homeostasis de las células madre hematopoyéticas en ratones (Siegemund et al., 2015). En pacientes con Alzheimer se ha detectado una sobreexpresión neuronal de esta isoforma que produce un incremento en muerte celular y producción de β -amiloides (Stygelbout et al., 2014).

En conclusión, la IP₃ 3-K posee una marcada relevancia biológica al desempeñar un papel importante en multitud de procesos celulares como la transducción de señales, el mantenimiento de la homeostasis del Ca²⁺, así como en el metabolismo de inositoles al ser el punto de partida de síntesis de otros IPs.

1.3.1.2 Inositol polifosfato multiquinasas (IPmKs)

1.3.1.2.1 Reacción catalítica

La enzima inositol polifosfato multiquinasas (IPmK) se encuentra ampliamente distribuida en levaduras, plantas y animales. Es una enzima nuclear y su pérdida de función es incompatible con la supervivencia (Resnick y Saiardi, 2008). Esta enzima constituye el miembro catalíticamente más diverso de esta familia ya que muestra actividad frente a más de un sustrato generalmente actuando como 3/5/6-quinasa. A través de su actividad 3-quinasa, convierte el I(1,3,4)P₃ en IP₄, bien sea el isómero I(1,3,4,5)P₄ o I(1,4,5,6)P₄, y posteriormente fosforila el IP₄ convirtiéndolo en I(1,3,4,5,6)IP₅ (Figura 1.10).

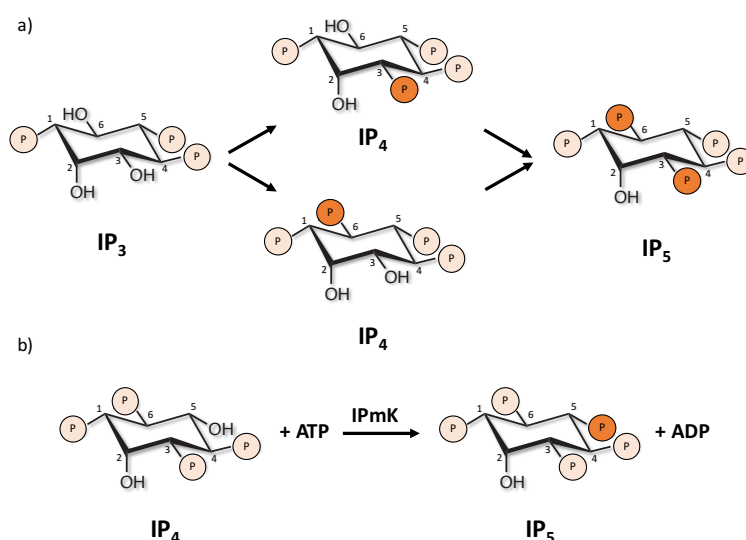


Figura 1.10. Principales reacciones catalizadas por la IPmK. a) Reacciones llevadas a cabo por la IPmK con actividad 3/6-quinasa. b) Reacción catalizada por la IPmK con actividad 5-quinasa.

La IPmK humana (y de planta) actúa principalmente como una 5-quinasa y es más activa como una 6-quinasa que como 3-quinasa. Por el contrario, en levaduras actúa preferiblemente como una 6- y 3-quinasa (Maag et al., 2011; Resnick et al., 2005; Saiardi et al., 2001).

1.3.1.2.2 Estructura de IPmK

La primera estructura de una IPmK se obtuvo en 2006 (código PDB 2IEW) (Holmes y Jogl, 2006), perteneciente a *Saccharomyces cerevisiae*. Posteriormente, en 2012 se obtuvo la estructura de *Arabidopsis thaliana* (código PDB 4FRF) (Endo-Streeter et al., 2012), siendo muy similar a la descrita en levadura. Muy recientemente, en septiembre de este mismo año se ha revelado la estructura de *Homo sapiens* (código PDB 5W2H) (Wang y Shears, 2017) (Figura 1.11).

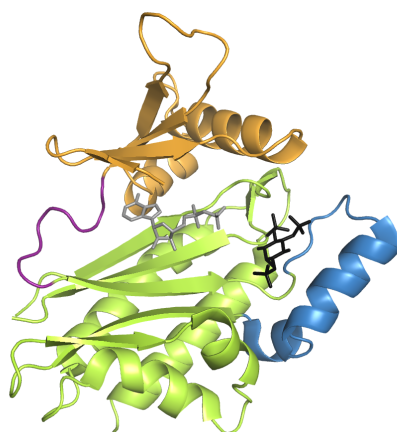


Figura 1.11. Estructura de IPmK de *Homo sapiens* (código PDB 5W2H). Diagrama de cintas siguiendo el código de colores anterior: en naranja se representa el lóbulo N-terminal, en verde el C-terminal y en azul el lóbulo de unión a IP. El ATP se muestra en gris, el IP₃ en negro y el conector en morado.

El plegamiento general de la IPmK es muy similar al de la IP₃ 3-K a pesar de que su identidad de secuencia es menor del 20%. Esta subfamilia tampoco presenta el lóbulo N completo tal y como se encuentra en las PKs, careciendo de alguna de sus hojas β . Como se ha mencionado anteriormente, el cambio más significativo entre estas dos enzimas ocurre en el sitio de unión a inositol, el cual es mucho menor en la IPmK ya que está formado por una única hélice larga, junto con otra muy pequeña. Esto se traduce en que el sitio activo de las IPmKs es mucho menos restringido y más abierto explicando su carácter altamente versátil a la hora de interactuar con distintos IPs, ya que estos se pueden situar en diferentes orientaciones. Seguramente, esta sea la causa de que hasta la estructura de IPmK de humano no se haya reportado ninguna isoforma en presencia de inositol. Por lo demás, los residuos de unión a ATP están conservados en posición y carácter, mientras que el par catalítico ácido/base están absolutamente conservados.

1.3.1.2.4 Funciones de la IPmK

Cabe destacar que la IPmK es una enzima multifuncional que posee actividad tanto inositol fosfato quinasa como actividad lípido quinasa. En cuanto a su actividad lípido quinasa, funciona como una PI3-quinasa catalizando la reacción de fosforilación en la posición 3 del PIP₂ para dar PIP₃ (ver esquema

global, Figura 1.2). El PIP₃ sirve como un activador fisiológico de la ruta akt/PKB (Maag et al., 2011; Resnick et al., 2005). Por otro lado, esta enzima posee funciones independientes de la catalítica. Así, la IPmK actúa como un cofactor fisiológico de mTOR y activador transcripcional para p53 induciendo la muerte celular tumoral (Xu et al., 2013). Otras dianas de IPmK son la proteína quinasa activada por AMP (AMPK) la cual es regulada fisiológicamente por IPmK en respuesta a la disponibilidad de glucosa (Bang et al., 2012; Dailey y Kim, 2012). La inactivación de IPmK disminuye la exportación nuclear de la proteína RAD51 lo que produce un descenso en la reparación del ADN (Wickramasinghe et al., 2013). Además, IPmK está implicada en una exportación selectiva de transcritos de ARNm, lo que hace preservar la integridad del genoma en humanos (Wickramasinghe et al., 2013). Finalmente, parece estar relacionada con los déficits patológicos y motores de la enfermedad de Huntington y con la presencia de tumores del tracto digestivo (de Mestier et al., 2017; Sei et al., 2015).

Todos estos hallazgos recientes evidencian que la IPmK juega un papel muy versátil en las redes de señalización celular.

1.3.1.3 Inositol pentaquisfosfato 2-quinazas (IP₅ 2-Ks)

1.3.1.3.1 Reacción catalítica e importancia general

La enzima inositol 1,3,4,5,6 pentaquisfosfato 2-quinasa (IP₅ 2-K) sintetiza el I(1,2,3,4,5,6)P₆ (IP₆) a partir del I(1,3,4,5,6)P₅ o IP₅ y ATP (Stephens et al., 1991) (Figura 1.12). Dentro de la familia de las IPKs, la IP₅ 2-K es la única enzima capaz de fosforilar la posición 2-OH axial del anillo de inositol, mientras que el resto de enzimas que pertenecen a la familia de las IPKs actúan sobre posiciones ecuatoriales. Es una enzima clave ya que la reacción de fosforilación que cataliza es crucial para la continuación del metabolismo de IPs (ver Figura 1.2). La IP₅ 2-K es una enzima muy específica, al igual que la IP₃ 3-K, ya que une un único sustrato y fosforila una posición única de los IPs, la posición 2 y 3 respectivamente (Phillippy et al., 1994).

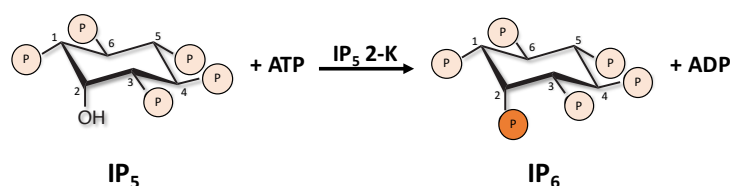


Figura 1.12. Reacción catalizada por la IP₅ 2-K. Síntesis de IP₆ a partir del IP₅ y ATP.

El producto de esta enzima, el IP₆, es el inositol más abundante en organismos vivos y juega un papel fundamental en procesos celulares, metabólicos y de señalización, que se describirá más adelante. En cuanto a la IP₅ 2-K, se ha demostrado que es una enzima clave en el desarrollo ya que la eliminación del gen (*Ipk1*) en ratones provoca una letalidad temprana en embriones (Verbsky et al., 2005). IP₅ 2-K también participa en el establecimiento de la asimetría de órganos izquierda-derecha en peces cebra (Sarmah et al., 2005).

1.3.1.3.2 Localización y funciones

IP₅ 2-K está presente en las células eucariotas, desde levaduras a mamíferos, aunque su identidad de secuencia está en el orden de sólo un 11% (Verbsky et al., 2002). El ortólogo de IP₅ 2-K en levaduras se denomina Ipk1. Es preciso señalar que no presenta homología de secuencia aparente con ninguna otra familia de enzimas, incluyendo otras familias de IPKs. En ratones adultos la IP₅ 2-K se expresa abundantemente en el hipocampo, en las células de Purkinge del cerebelo, en cardiomiocitos y en los testículos (Verbsky et al., 2005). En las células, IP₅ 2-K presenta una distribución dinámica, localizándose tanto en citoplasma como en el núcleo. Además, se ha demostrado que colocaliza con el ARNm bien en núcleo o citoplasma (Brehm et al., 2007). Recientemente se ha incrementado la atención prestada a esta enzima al demostrarse que presenta una doble funcionalidad, la catalítica y otra independiente de la misma.

En cuanto a la función catalítica y por tanto a la importancia de su producto IP₆, además de ser el precursor de los IPPs (IP₇ e IP₈) (Mulugu et al., 2007), este está involucrado en multitud de eventos celulares como exportación de ARNm (Montpetit et al., 2011; York et al., 1999), reparación del ADN (Cheung et al., 2008), regulación de la estructura de la cromatina (Hanakahi and West 2002) (Shen et al., 2003), apoptosis (Agarwal et al., 2009; Bozsik et al., 2007b), endocitosis (Efanov et al., 1997; Hoy et al., 2002), y regulación de canales iónicos (Yang et al., 2001). Algunos de los mecanismos moleculares que subyacen a estas funciones han sido determinados. Por ejemplo el IP₆ actúa como un cofactor estructural de enzimas de edición de ADN (Macbeth et al., 2005), induce cambios conformacionales en ADN-PK produciendo un aumento en la unión al ADN (Byrum et al., 2004), regula la actividad de la caseína quinasa 2 (CK2) mediante interacción directa con la subunidad α (Solyakov et al., 2004) y actúa como pegamento de complejos proteína-proteína que median la ubiquitinación del 20% del proteoma (Scherer et al., 2016). Además, existen varias evidencias que muestran beneficios de IP₆ como antioxidante o como potencial agente antitumoral (Graf et al., 1987) (Bozsik et al., 2007a). Específicamente en las plantas, el IP₆ o ácido fítico, se acumula en forma de sal (fítato) constituyendo la principal reserva de fósforo en las semillas, elemento necesario para su desarrollo y germinación (Raboy, 2003) y la resistencia basal a patógenos (Murphy et al., 2008). En contrapartida, el IP₆ puede ser causante de malnutrición y contaminación medioambiental. Por un lado, este es un potente quelante de iones esenciales como Zn^{+2} y Fe^{+2} y por tanto las dietas ricas en cereales contribuyen a la desnutrición (Hurrell, 2003). Por otro lado, los animales monogástricos son incapaces de digerir IP₆, lo que conduce a la eutrofización de las aguas. En este sentido, existe un elevado interés en la industria agropecuaria en el diseño de cosechas con bajo contenido de ácido fítico.

En cuanto a posibles funciones de IP₅ 2-K más allá de su función catalítica, entre otras, se ha reportado que presenta un papel estructural en el nucleolo de la célula relacionado con la síntesis de ARNr al interactuar con tres proteínas implicadas en la regulación de dicha síntesis: CK2, TCOF1 y el factor de unión “upstream” (UBF) (Brehm et al., 2007). En el apartado 1.4 se describen más funciones reportadas de IP₅ 2-K en relación con el metabolismo de ARN.

1.3.1.3.3 Estructura de IP₅ 2-K de planta, reconocimiento de sustrato y mecanismo

En 2010 se resolvió la primera estructura de IP₅ 2-K perteneciente a *Arabidopsis thaliana* (AtIP₅ 2-K) (Gonzalez et al., 2010) lo que permitió conocer que esta subfamilia es el miembro más distante dentro de la familia de las IPKs. Por un lado, AtIP₅ 2K tiene un lóbulo N que presenta las cinco hojas β de las

PKs, aunque incluyendo diferentes inserciones. No obstante, la mayor diferencia con las IPKs radica en el lóbulo C, ya que presenta insertada una región estructural singular para la unión de inositol que abarca casi la mitad de la proteína. Este lóbulo IP está formado por tres segmentos no continuos en secuencia (Figura 1.13a). Otra gran novedad es que *AtIP*₅ 2-K es una zinc-metalo enzima. Todos los residuos que forman el sitio de unión a zinc están localizados en uno de los segmentos del lóbulo IP (Figura 1.13b).

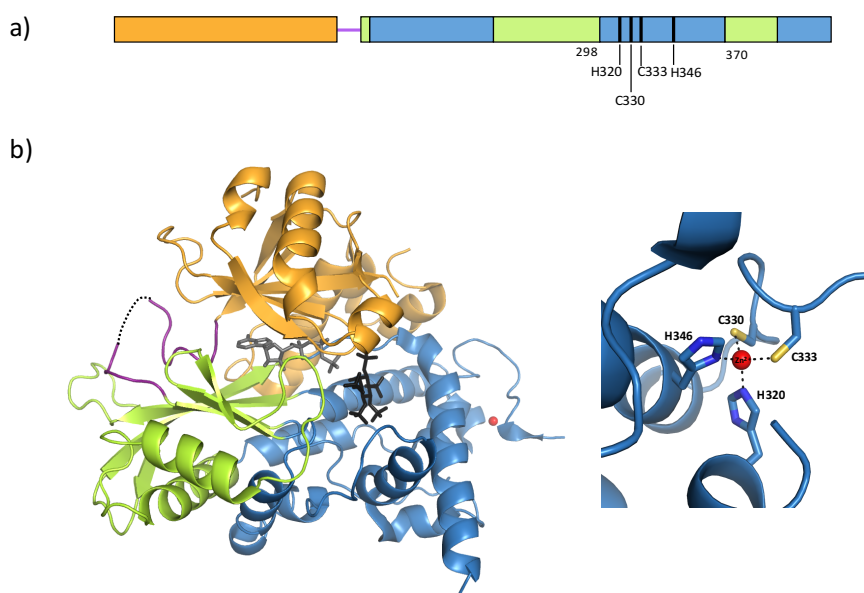


Figura 1.13. Estructura de IP₅ 2-K de *Arabidopsis thaliana* (PDB 2XAN). a) Esquema de los dominios de IP₅ 2-K, en naranja se representa el lóbulo N-terminal, en verde el C-terminal y en azul el lóbulo de unión a IP. b) A la izquierda, representación de la estructura mediante diagrama de cintas siguiendo el código de colores establecido. El ATP se muestra en gris, el IP₅ en negro, en rojo el zinc y el conector en morado. A la derecha, detalle del sitio de unión a zinc y residuos que lo coordinan.

El bolsillo de unión al sustrato IP₅ está formado en su mayoría por residuos del lóbulo C (Figura 1.14). Sin embargo, en esta familia, el lóbulo N también participa en el reconocimiento de IP₅, lo cual es fundamental para terminar de completar el sitio de unión a inositido. En particular, el lóbulo N reconoce los fosfatos P1 (Lys130) y P3 (Arg45 y bucle-G). *AtIP*₅ 2-K también conserva el par ácido/base catalítico (Asp407/Lys168), presentando dos elementos nuevos y característicos de su familia: un segundo residuo (Ser409) que coordina el Mg²⁺ junto con el ácido anterior y un ácido (Asp368) que coordina directamente el 2-OH del sustrato IP₅. El papel de este ácido es fijar la posición del 2-OH. Todas y cada una de las múltiples interacciones del IP₅ con la enzima orientan a este sustrato de manera que el anillo de inositol permanece perpendicular con respecto a la orientación encontrada en otras IPKs. Esta orientación permite mantener una geometría de reacción adecuada para que sea posible la fosforilación de la posición 2 del anillo (Figura 1.13).

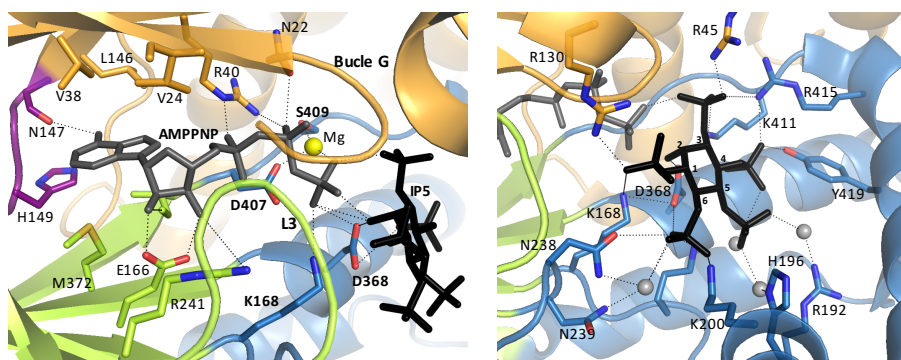


Figura 1.14. Centro activo de la IP₅ 2-K. Izquierda, detalle de los residuos implicados en el reconocimiento del AMPPNP (gris) y derecha, del IP₅ (negro). En negrita se resaltan los residuos catalíticos Asp407, Lys168, Ser409 y Asp368. Las aguas se representan en esferas grises y el ión Mg²⁺ en amarillo.

Estudios estructurales han mostrado que existe un cambio conformacional asociado a la catálisis (Banos-Sanz et al., 2012). La unión del ATP y el inositol producen ciertos cambios conformacionales; así en la forma no unida la IP₅ 2-K muestra una conformación abierta, mientras que si está unida al nucleótido muestra una conformación semi-cerrada, en contraste con la forma unida a inositol que es cerrada. Estos cambios pueden entenderse como una aproximación del lóbulo N-, una vez que el inositol se une en el sitio de unión a IP, para completar esta unión (Figura 1.15).

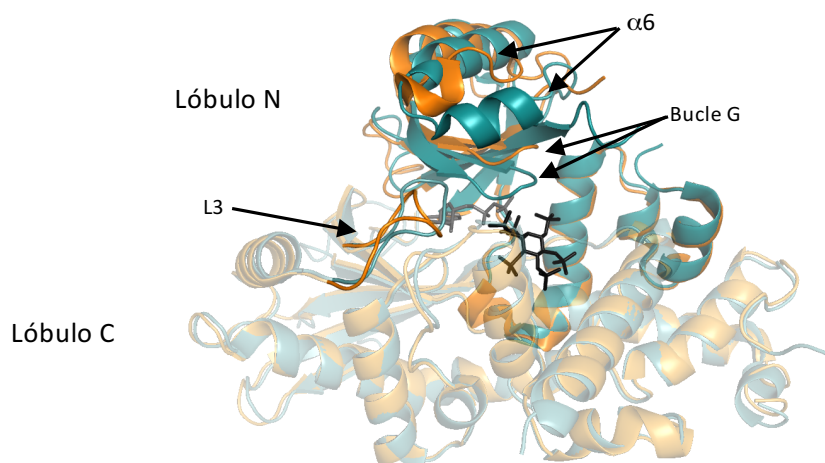


Figura 1.15. Cambio conformacional en IP₅ 2-K de *Arabidopsis thaliana*. Superposición de las estructuras de IP₅ 2-K en ausencia/presencia de inositido (códigos PDBs 2XAL/2XAM). En naranja se representa la IP₅ 2-K en su forma apo y en azul el complejo ternario de IP₅ 2-K/IP₆/ADP. El ADP se representa en gris y el IP₆ en negro. Con flechas negras se destacan tres regiones afectadas por los cambios conformacionales: bucle 3 (L3), bucle G y hélice α6. El lóbulo C se representa en transparencia.

Las diversas estructuras obtenidas de *AtIP*₅ 2-K en su forma apo y en presencia nucleótidos, IPs y combinación de ambos han permitido conocer los elementos estructurales clave en el cambio de conformación. En el lóbulo N-terminal la hélice α6, portadora de la Arg130 implicada en la unión del P1, se estructura al unir el inositol. Por otro lado, el bucle G es altamente flexible y se adapta para unir los fosfatos del ATP y el P3 del IP₅. Aunque el lóbulo C permanece sin cambios globales, cabe destacar el cambio de conformación producido en el bucle L3, el cual cambia el esquema de interacción con diferentes residuos implicados en la unión a inositol. Posteriormente a este trabajo, y en consonancia

con el mismo, varios estudios bioquímicos apoyaron este mecanismo de reconocimiento del sustrato, demostrando que la unión del inositol juega un papel fundamental en la estabilización del lóbulo N-terminal y posterior activación de la enzima (Gosein et al., 2012; Gosein y Miller, 2013a, b).

En conclusión, la IP₅ 2-K además de ser una enzima clave dentro de la compleja red de IPs, está involucrada en multitud de procesos celulares e interacciones con otras proteínas que ponen de manifiesto su importancia. Todos estos hallazgos aumentan el potencial de esta enzima como una diana atractiva para su estudio.

1.3.1.4 Inositol hexakisfosfato quinasas (IP₆Ks)

1.3.1.4.1 Reacción catalítica e IPPs

En cuanto a su actividad catalítica, las IP₆Ks poseen actividad 5-quinasa que da como resultado los inositoles pirofosfatos (IPPs): 5-IP₇ y el IP₈. Las IPKs transfieren un grupo fosfato desde el ATP a la posición 5 del IP₆ sintetizando el 5-IP₇ (5-bisfosfoinositol pentaquisfosfato) un isómero de IP₇ que tiene dos grupos fosfatos en la posición 5 de anillo de inositol. Adicionalmente, fosforilan el isómero 1-IP₇ en la posición 5 del anillo dando lugar al IP₈ (1,5-bisfosfoinositol tetraquisfosfato) que tiene dos grupos fosfato en las posiciones 1 y 5 (Saiardi, 2012). Las IP₆Ks colaboran junto con las PPIP5K (quinasas con plegamiento “ATP-grasp”, más información en el apartado 1.3) en la síntesis de los IPPs a través de dos vías metabólicas complementarias (Figura 1.16). Además de la participación de las IP₆Ks en la síntesis de los IPPs, están involucradas en la fosforilación del I(1,3,4,5,6)P₅ a 5-PP-IP₄ (Mulugu et al., 2007) (Figura 1.16).

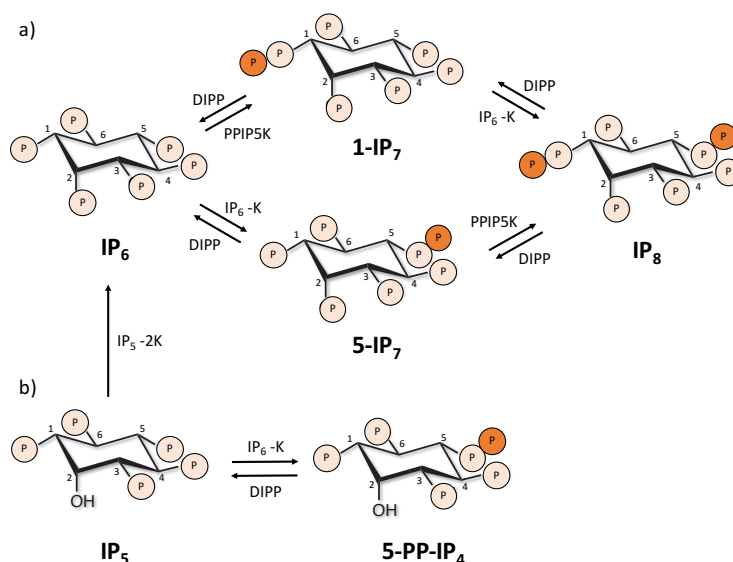


Figura 1.16. Reacciones de formación IPPs. a) Rutas complementarias de síntesis de los IPPs a través de las IP₆Ks e PPIP5K. b) Ruta de fosforilación de 5-PP-IP₄ a partir de IP₅.

Los IPPs presentan una localización ubicua en las células eucariotas y son sintetizados por dos familias distintas que exhiben actividad catalítica frente a diferentes posiciones en el anillo, una de ellas la IP₆K (Wilson et al., 2013). A su vez estos compuestos pueden ser degradados por la familia de fosfatasas

difosfo inositol polifosfato fosfohidrolasa (DIPP) (Safrany et al., 1999). Los IPPs desempeñan un papel importante en muchos procesos celulares incluyendo el mantenimiento de la longitud de los telómeros (Saiardi et al., 2005; York et al., 2005), tráfico de vesículas y regulación de la morfología celular en levaduras (Saiardi et al., 2000), la autofagia (Nagata et al., 2010), la reparación del ADN mediante recombinación homóloga en levaduras (Jadav et al., 2013), la apoptosis (Morrison et al., 2001), la secreción y señalización por insulina (Chakraborty et al., 2010; Illies et al., 2007), exocitosis, quimiotaxis (Luo et al., 2003), desarrollo embrionario y regulación del ciclo celular (Lee et al., 2007). Después de más de dos décadas de trabajo, se comienza ahora a tener una visión de los IPPs como agentes fundamentales, conservados y de gran importancia en la regulación de la homeostasis celular (Szijgyarto et al., 2011; Wilson et al., 2013).

1.3.1.4.2 Isoformas de IP₆K y funciones

Los mamíferos poseen tres isoformas de la IP₆Ks, IP₆K1, IP₆K2 e IP₆K3, las cuales presentan una identidad en secuencia en torno al 45%, y su máxima diferenciación radica en una larga inserción (>100 residuos) que presentan en el lóbulo N-terminal; además, presentan una inserción menor de 45 residuos en el lóbulo C-terminal. Mientras que las IP₆K1 y IP₆K2 son el producto de los genes IP₆K1 e IP₆K2 respectivamente, localizados en el cromosoma 3, la IP₆K3 es el producto del gen IP₆K3 localizado en el cromosoma 6 humano. Los productos de estos genes codifican para unas proteínas de 50 kDa, 49 kDa y 46 kDa respectivamente. Mientras que IP₆K1 e IP₆K2 se localizan tanto en núcleo como citoplasma, la IP₆K3 es exclusiva del citoplasma. Las isoformas IP₆K1 e IP₆K2 se expresan en la mayoría de tejidos mientras que la IP₆K3 es única de las células de Purkinje del cerebelo (Saiardi et al., 2001). La principal expresión de IP₆K1 se localiza en testículos y la IP₆K2 en pulmón (Mulugu et al., 2007). Se ha demostrado que las tres isoformas participan en funciones celulares no superpuestas como se explica en la próxima sección. En relación al sustrato, todas las isoformas presentan una preferencia por el IP₆, si bien la IP₆K2 muestra la mayor afinidad de las tres.

La función de las diferentes isoformas no es redundante. De este modo, IP₆K1 regula numerosas funciones fisiológicas, como la secreción de insulina por las células pancreáticas (Bhandari et al., 2008; Chakraborty et al., 2010), la reparación del ADN mediada por recombinación homóloga y la formación de espermatozoides maduros (Jadav et al., 2013). Se ha postulado que posibles inhibidores de IP₆K1 podrían ser agentes terapéuticos útiles en la diabetes (Koldobskiy y Snyder, 2011). Su eliminación produce malformación cerebral durante el desarrollo neural (Fu et al., 2017). Por otro lado, ratones que carecen de IP₆K2 muestran desarrollo normal, crecimiento y fertilidad, pero están predispuestos al desarrollo de carcinoma del tracto aerodigestivo (Morrison et al., 2009). En relación a la enfermedad de Huntington, la IP₆K2 promueve la muerte celular (Nagata et al., 2011). Un estudio reciente ha propuesto que la IP₆K2 podría servir como un biomarcador para el diagnóstico temprano de la ELA (esclerosis lateral amiotrófica) antes de la aparición de los trastornos motores que conlleva esta enfermedad (Yu et al., 2016). Finalmente, ratones “*knock out*” para IP₆K3 muestran disfunción cerebral al presentar una alteración en la formación sináptica además de una reducción en la coordinación y el aprendizaje motor (Fu et al., 2015). Recientemente se ha descrito que cierta variabilidad en el gen de la IP₆K3 podría estar relacionada con la patogénesis de la enfermedad de Alzheimer (Crocco et al., 2016). Por último, se ha observado que posee un papel fundamental en la regulación del control metabólico y la esperanza de vida media en ratones (Moritoh et al., 2016).

1.3.1.4.3 Interacciones proteína-proteína

Al igual que el resto de las IPKs, las funciones de las IP₆Ks derivan de su actividad catalítica o de sus interacciones con otras proteínas. Algunas de estas interacciones ya se han descrito. En particular, IP₆K1, mediante su extremo N-terminal, se asocia con un factor de intercambio de nucleótidos de guanina (GEF) denominado GRAB que promueve la liberación de células neuronales (Luo et al., 2001). Hasta la fecha, se han identificado tres proteínas que interaccionan con IP₆K2: el factor asociado al receptor de necrosis tumoral TRAF2, la proteína de choque térmico HSP90 y p53 (Koldobskiy et al., 2010). Por un lado, el extremo C-terminal de HSP90 se une fisiológicamente al extremo N-terminal de IP₆K2 e inhibe su actividad catalítica produciendo una reducción en los niveles de los IPPs (Chakraborty et al., 2008; Shames y Minna, 2008). Por otro lado, la unión a TRAF2 mediante la región C-terminal es independiente de su actividad quinasa (Morrison et al., 2007). Muchos trabajos sugieren que la unión de IP₆K2 a p53 ejerce un papel clave en apoptosis (Koldobskiy et al., 2010; Nagata et al., 2005). Finalmente, la CK2 promueve la supervivencia celular y metástasis de tumores al fosforilar a IP₆K2 produciendo su ubiquitinación y posterior degradación (Chakraborty et al., 2011) (Figura 1.17).

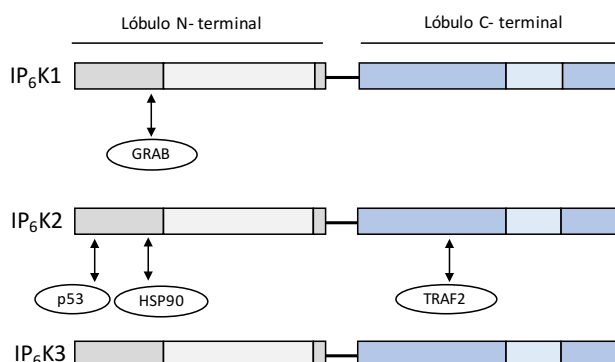


Figura 1.17. Organización general de las IP₆Ks. Esquema de las isoformas de IP₆K mostrando en gris oscuro y azul oscuro las zonas más conservadas entre las isoformas, y en claro las regiones más variables: en gris claro la inserción de 100 residuos en el lóbulo N-terminal y en azul claro C la inserción de 45 residuos del C-terminal. Se representan también las distintas proteínas con las que interacciona cada isoforma y las regiones de interacción.

1.3.1.4.3 Estructura de IP₆K en protozoos

En 2014 se determinó la única estructura conocida hasta la fecha de una IP₆K, si bien se trata de la enzima de un organismo protozoo (*Entamoeba histolytica*) (Wang et al., 2014a) (Figura 1.18). Esta IP₆K presenta una secuencia muy distinta a la IP₆K de mamífero y carece de la llamativa inserción en el lóbulo N- anteriormente mencionada. Además, esta proteína a diferencia de las demás IP₆Ks, posee una actividad bifuncional 3/6-quinasa. La estructura reveló que comparte elementos estructurales con IPmKs e IP₃ 3-K a la vez que revela diferencias que contribuyen a la actividad catalítica específica. Entre los elementos conservados están el dominio N y C-terminal que forman el sitio de unión a ATP. El dominio de unión a IP está formado por una única hélice seguida de otra muy pequeña igual que en el caso de las IPmKs, configurando un bolsillo de unión a IP₆ abierto y relativamente espacioso.

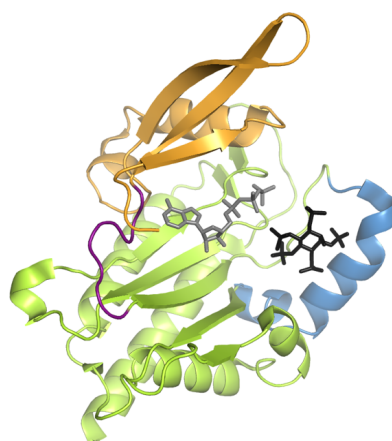


Figura 1.18. Estructura de IP₆K de *Entamoeba histolytica* (código PDB 4O4C). Representación estructural de IP₆K mediante diagrama de cintas. En color naranja se representa el lóbulo N-terminal, en verde el C-terminal y en azul el lóbulo IP. El ATP se muestra en gris, el IP₆ en negro y el conector en morado.

En conclusión, la familia de IP₆K posee una actividad dual, con actividad quinasa relacionada con el metabolismo de inositoles y, por otro lado, múltiples funciones derivadas de sus interacciones proteína-proteína.

1.4 Interferencias entre las rutas de IPs y biogénesis del ARN

A lo largo de esta introducción ha quedado reflejado que muchas de las IPKs y las funciones de los IPs se basan en la interferencia y/o interacción directa o indirecta con diferentes mecanismos que intervienen en la biogénesis del ARN. En concreto, se ha visto relación entre los IPs con la transcripción y su regulación, edición y mecanismos de vigilancia del ARN y exportación del ARN. A continuación, se detalla brevemente algunos de estos papeles de la ruta de IPs:

1.4.1. Ruta de IPs y regulación de la transcripción

La información genética se encuentra en grupos de secuencias codificantes y no codificantes. En eucariotas esta información se transcribe por tres ARN polimerasas (ARN Pol I, ARN Pol II y ARN Pol III), cada una especializada en un tipo de transcrito. La ARN Pol I sintetiza el ARNr, la ARN Pol II sintetiza ARNm, microARNs y ARNs no codificantes, y, por último, la ARN Pol III sintetiza el ARNt y otros ARN pequeños. Se han encontrado múltiples relaciones entre la ruta de los IPs (Figura 1.2) y la transcripción mediada por las ARN polimerasas:

- Una serie de estudios sugieren que tanto la IP₃ 3-K como la IPmK intervienen en la regulación de la actividad transcripcional. En levaduras, estas enzimas funcionan como componentes reguladores del complejo transcripcional ArgR-Mcm1, requerido para la supresión/inducción de varios productos génicos que median catabolismo y síntesis de arginina (Odom et al., 2000; Resnick et al., 2005).
- Hallazgos recientes relacionan los IPPs con la regulación de la síntesis de ribosomas y se les designa como nuevos reguladores de la transcripción del ARNr en *Saccharomyces cerevisiae* mediante la regulación de la ARN Pol I (Thota et al., 2015).

- Se ha demostrado la participación de la enzima IP₅ 2-K en la biogénesis del ARNr al interactuar con las proteínas CK2 y UBF, ambas implicadas en la estimulación de la ARN Pol I (Brehm et al., 2007).

- Se sabe que la señalización mediante RAS promueve la transcripción de los genes implicados en el crecimiento y glucólisis. En levaduras, la ruta Nrd1-Nab3 parece tener un papel antagónico al inducir una supresión rápida de algunos genes cuando las células se encuentran en condiciones de crecimiento deficientes/baja disponibilidad de nutrientes (Darby et al., 2012). Si bien no se conoce el mecanismo implicado en esta supresión, si se conoce la función de Nrd1-Nab3 en la terminación de la transcripción de transcritos no codificantes generados por la ARN Pol II. Para ello, Nrd1-Nab3 reconocen secuencias cortas ARN de cadena sencilla. Por otro lado, Nrd1 es capaz de reclutar al exosoma (complejo multiproteico capaz de procesar o degradar diversos tipos de ARNs) dirigiendo el ARN para su procesamiento o degradación (Vasiljeva y Buratowski, 2006). Por el momento no se conoce con precisión cómo estas funciones de Nrd1 están relacionada con su papel antagónico a la señalización por RAS.

1.4.2. IP₆ y edición del ARN

Como se ha mencionado anteriormente, el IP₆ actúa como un cofactor estructural para diferentes enzimas de edición de ARN. Este es el caso de las adenosinas desaminasas ADAR2 y ADAT1 que participan en la edición del ARNm y ARNt respectivamente. Estas enzimas catalizan la edición de ARN desaminando la adenosina a inosina en el ARN de doble cadena, participando en la regulación de la expresión génica a nivel del ARN. En particular, el IP₆ se une en el núcleo del dominio catalítico, contribuyendo tanto en el plegamiento y estabilidad de la proteína, como en la actividad enzimática (Macbeth et al., 2005).

1.4.3. IP₅ 2-K en la vigilancia del ARN

Los mecanismos de vigilancia del ARNm son vías utilizadas para detectar y degradar los transcritos que no se han procesado correctamente y garantizar así la fidelidad y la calidad de las moléculas de ARNm. Existe varias vías de vigilancia del ARNm, una de ellas es la denominada “*nonstop mRNA decay*” relacionada con el exosoma. Mediante esta vía se identifican y degradan transcritos aberrantes que codifican para proteínas que podrían causar efectos dañinos en la célula. En concreto, se detectan moléculas de ARNm que carecen de un codón de parada evitando así que estos ARNm se traduzcan. Mediante métodos genómicos en levaduras, se ha detectado que la IP₅ 2-K está implicada en la inhibición de esta ruta (Wilson et al., 2007).

1.4.4. IP₆K e IPmK en exportación del ARN

En los últimos años, ha quedado reflejada la importancia del IP₆ en la exportación del ARNm. En concreto, la proteína Gle1 estimula la función de la helicasa de ARN dependiente de ATP (Dbp5), una proteína que se une al ARNm para dirigir su exportación desde los núcleos. Pero esta estimulación solamente se produce cuando Gle1 está unido a su cofactor IP₆ (Montpetit et al., 2011). Por otro lado, también se ha reportado el papel clave de la IPmK en la exportación del ARNm desde el núcleo hasta el

citoplasma. Esta promueve la exportación de ARNm nucleares específicos como por ejemplo el ARNm de RAD51 que codifican para proteínas de recombinación sensibles al daño oxidativo (Kim et al., 2017; Kim et al., 2016; Wickramasinghe et al., 2013).

Podemos observar que muchas de las conexiones entre los dos metabolismos (IPs y ARN) se producen mediante interacciones proteína-proteína o proteína-IP, en muchos casos regulando a proteínas que interaccionan directamente con el ARN. Al día de hoy, disponemos de un gran conocimiento estructural y del mecanismo de acción de muchas de estas proteínas (Gle1, ADAR2, CK2) o alguno de sus dominios (Nab3, ARN Pol, UBF), si bien en otros casos hay menos información. En este trabajo se ha llevado a cabo el estudio estructural del dominio de reconocimiento de ARN de la proteína Nrd1 y su reconocimiento de ARN.

1.5 Reconocimiento proteína-ARN: Nrd1 y papel en transcripción

1.5.1 Motivos de reconocimiento de ARN: dominios “RRM”

Las interacciones entre proteínas y ARN son esenciales para la biogénesis, control y función del ARN. Existen diversos tipos de dominios de unión a ARN, siendo los motivos de reconocimiento de ARN (“*RNA Recognition Motif*” o “RRM”) unos de los más abundantes (Maris et al., 2005). Los RRM unen ARN de cadena sencilla con un rango de especificidades y afinidades muy amplio. Estos están presentes mayoritariamente en células eucariotas, aunque también se encuentran en procariotas y virus. Participan en multitud de eventos post-transcripcionales de la célula, mostrando gran variedad de funciones celulares como el “*splicing*” alternativo, la estabilidad y procesamiento del ARN, la regulación de la traducción, exportación y degradación del ARN (Maris et al., 2005). Los RRM son dominios pequeños de unos 90 aminoácidos con una arquitectura sencilla, aunque pueden presentar una gran versatilidad estructural ya que puede contener multitud de modificaciones que varían su afinidad y selectividad por el ARN. La estructura canónica típica de un dominio RRM incluye una arquitectura tipo $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$, en la que las dos hélices α se encuentran empaquetadas contra la hoja β antiparalela formada por cuatro hebras β (Figura 1.21).

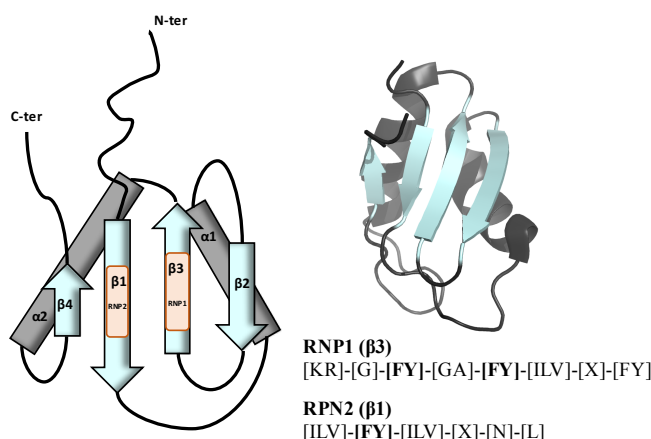


Figura 1.21. Estructura canónica de los dominios de unión a ARN tipo RRM. Esquema de la topología del dominio RRM (izquierda) y representación estructural como diagrama de cintas (derecha) de la proteína Nab3 (código PDB 2KVI). Las cuatro hebras β antiparalelas que forman la hoja β se muestran en cian mientras que las hélices α se

muestran en gris. El esquema de la parte inferior muestra las secuencias conservadas RNP1 y RNP2 en las hebras $\beta 3$ y $\beta 1$ respectivamente.

Dentro del dominio RRM hay dos motivos centrales muy conservados en secuencia denominados RNP (“*Ribonucleoprotein domain*”). Las hebras $\beta 3$ y $\beta 1$ contienen dos motivos: RNP1 formado por un segmento conservado de 8 residuos, y un segmento de 6 residuos menos conservado, RNP2. En estas dos secuencias se encuentran los tres aminoácidos aromáticos (la posición 2 de la RNP2 y las posiciones 3 y 5 de la RNP1 resaltadas en negrita en la secuencia de la Figura 1.21 que refuerzan la interacción con ARN mediante un apilamiento con sus bases nitrogenadas (Clery et al., 2008; Maris et al., 2005).

1.5.2 Nrd1 y terminación de la transcripción

El fenómeno de la transcripción es el primer proceso de la expresión génica y en él se transfiere la información contenida en la secuencia del ADN hacia la secuencia de proteínas utilizando diversos ARNs como intermediarios. La información genética se encuentra en módulos (genes) codificantes y no codificantes. En levaduras, la ARN Pol II (Pol II) transcribe no solo genes codificantes de proteínas (típicamente ARNm), sino también ARNs pequeños en general no codificantes (ARNnc) como el ARN pequeño nuclear (ARNsn) y ARN pequeño nucleolar (ARNsno), y más recientemente se han descrito los transcritos no codificantes de función menos conocida como transcritos crípticos inestables (“*Cryptic Unstable Transcripts*” o “CUTs”) y transcritos estables sin anotar (“*Stable Unannotated Transcripts*” o “SUTs”). Los ARNnc regulan diversos procesos celulares a través del control de la expresión génica. De este modo participan en una amplia variedad de procesos celulares como la regulación de la replicación del ADN, la transcripción, traducción y el ensamblaje o *splicing* del ARN (Yamamura et al., 2017). Además, están implicados en procesos patológicos como el cáncer (Naeini y Ardekani, 2009) y enfermedades neurodegenerativas como el Alzheimer, el Parkinson, la enfermedad de Huntington y la ELA (Quan et al., 2017).

El paso final de terminación de la transcripción en las levaduras se lleva a cabo a través de dos vías dependiendo de la naturaleza del transcrito. De esta manera, mientras que la terminación de la transcripción en el ARNm es terminada por un complejo dependiente de poliadenilación (poli(A)), la terminación de la transcripción del ARN no codificante es dependiente del complejo proteico denominado NNS independiente de poliadenilación. La elección de una u otra ruta depende de la presencia de motivos de secuencia específicos en la región 3’UTR de los ARNs y, sorprendentemente, también por su longitud. Así el complejo NNS se recluta en etapas tempranas de la transcripción asociado a la marca CTD S5p del complejo de ARN Pol II, mientras que el complejo dependiente de poliadenilación se recluta de forma tardía asociada a la marca CTD S2p. El complejo NNS está formado por tres proteínas, dos de ellas están involucradas en el reconocimiento específico de señales en el ARN mediante dominios RRM: Nrd1 y Nab3 y, la tercera proteína es la ARN helicasa Sen1 (Figura 1.22).

Nrd1 es una proteína modular que en su extremo N-terminal posee el dominio CID (“*CTD Interaction Domain*”), en la región central un dominio de heterodimerización con Nab3 denominado DD, y en el extremo C-terminal el dominio RRM (Figura 1.23).

A través del dominio CID, Nrd1 interacciona con dos elementos; por un lado, con el dominio CTD (“*C-terminal Domain*”) de la ARN Pol II en un estadio temprano de la transcripción para terminar los transcritos cortos no codificantes; por otro lado, con la proteína Trf4 del complejo multiproteico

TRAMP, que dirige los transcritos al exosoma nuclear para su posterior degradación (Tudek et al., 2014) (Figura 1.22). Se conoce la estructura del CID de Nrd1, formado por ocho hélices α (Vasiljeva y Buratowski, 2006) (Figura 1.23), tanto aislado como en complejo con el CTD de la ARN Pol II (Kubicek et al., 2012; Vasiljeva y Buratowski, 2006) así como en complejo con Trf4 (Tudek et al., 2014). El dominio CID interacciona con el CTD y Trf4 de una manera mutuamente exclusiva, definiendo así dos formas alternativas del complejo NNS, una involucrada en la terminación y la otra en degradación. En relación al dominio de dimerización, Nrd1 forma un heterodímero con Nab3 a través de este dominio, del que no hay información estructural. En último lugar, en 2014 se determinó la estructura parcialmente desestructurada del dominio RRM de Nrd1 mediante RMN (código PDB 2M88) (Bacikova et al., 2014) (Figura 1.23). Sin embargo, este trabajo no revela el plegamiento que presenta el dominio de unión a ARN completo, el cual se extiende más allá de la secuencia canónica del RRM. De la misma forma, no se aporta ningún detalle del reconocimiento de ARN por parte de Nrd1, si bien se conoce que la secuencia de reconocimiento es GUA(A/G) gracias a numerosos estudios del transcriptoma de levadura.

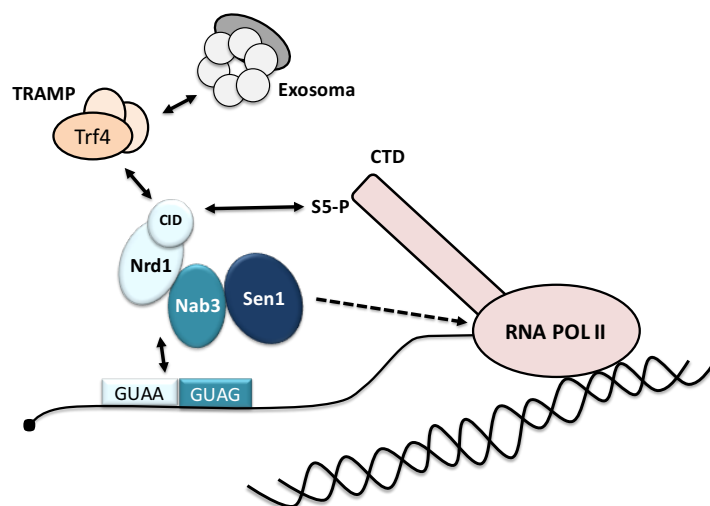


Figura 1.22. Interacción del complejo NNS con ARN Pol II y el complejo TRAMP. Nrd1 es capaz de interactuar a través de su dominio CID con el CTD de la ARN Pol II a través de la serina 5 fosforilada (S5-P), y con Trf4 del complejo TRAMP que a su vez interactúa con el exosoma. Nrd1 además se une al ARN naciente que contiene las secuencias GUAA.

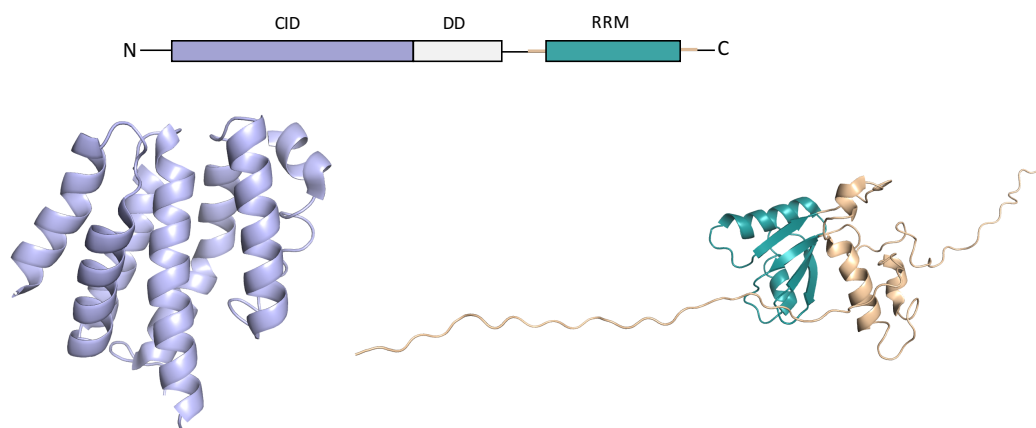


Figura 1.23. Esquema de la Organización modular de Nrd1. Nrd1 contiene en su extremo N-terminal el dominio CID, un dominio de dimerización (DD) en la parte central y a continuación el dominio RRM. Abajo a la izquierda, en azul añil estructura del dominio CID (código PDB 3CLJ). Abajo a la derecha, estructura parcial del dominio de unión a ARN de Nrd1 (código PDB 2M88), en verde azulado se muestra el RRM y en crema la estructura de las regiones que rodean al RRM.

1.5.3. Otros componentes del complejo NNS

Nrd1 forma complejo con las proteínas Nab3 y Sen1, cuyas estructuras completas o parciales ya se conocen. En particular, Nab3 reconoce, junto con Nrd1, elementos específicos de la terminación del ARN naciente, UCUU y UCUG en este caso (Carroll et al., 2004; Porrúa et al., 2012) que se encuentran frecuentemente en secuencias de terminación de los ARNnc (Figura 1.22). Se conoce la estructura del RRM de Nab3, tanto aislado (código PDB 2KVI) como en complejo con el ARN UCUU (código PDB 2L41) (Hobor et al., 2011) (código PDB 2XNR) (Lunde et al., 2011) (Figura 1.24a). Las especificidades de unión demostraron que el RRM de Nab3 establece una clara preferencia por la citidina central, aunque las uridinas flanqueantes también son importantes para la unión. Por otro lado, Sen1 posee una actividad helicasa ARN-ADN y muy recientemente, en 2017 y también mediante Cristalografía de rayos X, se ha resuelto su estructura (código PDB 5MZN) (Leonaite et al., 2017) (Figura 1.24b), lo que ha permitido comprender cómo esta enzima lleva a cabo la rotura de los puentes de hidrógeno y el desenrollado del ARN-ADN.

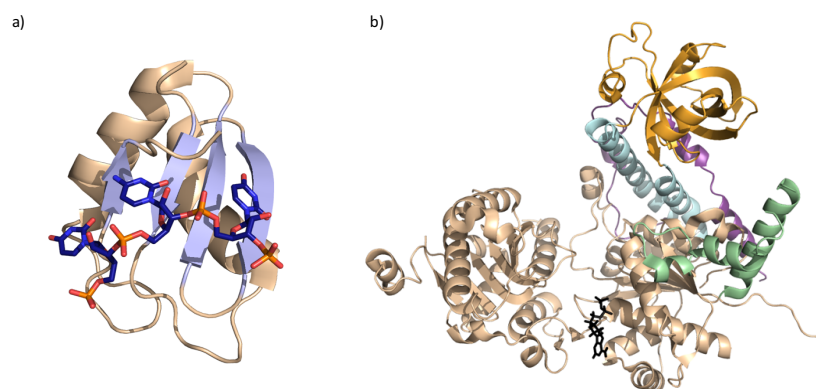


Figura 1.24. Estructuras de Nab3 y Sen1. a) Estructura del complejo formado por el dominio RRM de Nab3 con UCUU (código PDB 2XNR). En azul oscuro se muestra el ARN UCUU (con los oxígenos en rojo y los fosfatos en naranja), en azul claro las hojas β y en crema las hélices α del RRM. b) Estructura de la helicasa Sen1 (código PDB 5MNZ), que muestra la complejidad de esta proteína.

Como se puede observar se dispone de bastantes datos estructurales del complejo NNS, lo cual es crucial para entender este complejo sistema. Sin embargo, todavía existen muchas incógnitas como la estructura del dominio íntegro de unión a ARN de Nrd1, cómo esta proteína reconoce el ARN, y cuáles de sus mecanismos de actuación implican una función contraria a la señalización mediante la proteína RAS.

En conclusión, actualmente existen multitud de datos estructurales de todas las subfamilias de IPKs, especialmente en plantas y levaduras u organismos más elementales. Sin embargo, y a pesar de todas las funciones que presentan, aún quedan por descifrar las estructuras de las isoformas de mamífero de IP₅ 2-

K e IP_6K . La determinación de las estructuras de estas isoformas es clave para entender su especificidad por los diferentes IPs y sus mecanismos catalíticos y de regulación. Además, estos datos podrían ayudar a elucidar su posible interacción con otras proteínas y el entendimiento de funciones diferentes a la catalítica. En cuanto a la proteína Nrd1, se carece de información estructural completa sobre el dominio de unión a ARN y sobre el reconocimiento del mismo. La presente tesis doctoral recoge el estudio estructural de todas estas proteínas con el fin de esclarecer etapas clave en el metabolismo de los IPs y del ARN.

2. OBJETIVOS

2. Objetivos

El objetivo principal de esta tesis es llevar a cabo estudios de biología estructural enmarcados en dos líneas de investigación que engloban proteínas implicadas en la regulación de la ruta de los IPs y su relación con el ARN.

Por un lado, como se ha expuesto en el apartado anterior, la red de señalización celular por IPs es clave en diferentes procesos biológicos. Obtener información sobre esta compleja red desde un punto de vista estructural es crucial para entender mejor su funcionamiento y, en consecuencia, contribuir a la resolución de los problemas de salud que subyacen en una señalización incorrecta. Aunque en los últimos años se ha hecho un gran avance en el estudio de las IPKs, esta tesis tiene como objetivo general profundizar en el estudio de la biología estructural de tres enzimas de esta familia. Los objetivos concretos son:

- 1. Obtención de la estructura de IP₅ 2-K de mamífero.** Desarrollo de protocolos de expresión y purificación de la proteína IP₅ 2-K de mamífero (proteína de ratón o *mIP₅ 2-K*). Cristalización y resolución estructural de *mIP₅ 2-K* mediante Cristalografía de rayos X. Caracterización estructural de *mIP₅ 2-K* en presencia de sustratos e identificación de los residuos implicados en la especificidad de unión al sustrato. Análisis de las características específicas de esta familia en mamíferos y comparación con la isoforma de planta. Diseño de experimentos de mutagénesis dirigida y biofísicos para apoyar los resultados estructurales. Identificación de nuevas regiones implicadas en la función biológica de IP₅ 2-K en mamíferos.
- 2. Producción y estudio estructural de la IP₆K de mamífero.** Puesta a punto de protocolos de expresión y purificación de las diferentes isoformas de IP₆K de humano. Abordaje de experimentos de cristalización, utilización y desarrollo de estrategias para la cristalización de muestras difíciles. Obtención de información estructural relevante para comprender la función de esta familia.

Por otro lado, en la sección anterior ha quedado reflejado que los IPs y sus enzimas relacionadas interfieren frecuentemente con rutas relacionadas con la biogénesis y metabolismo del ARN. Otro de los objetivos de este trabajo es entender y explorar esta relación y el reconocimiento de ARN mediante proteínas. Para ello, se pretende abordar el estudio cristalográfico de la proteína Nrd1 implicada en la terminación de la transcripción:

- 3. Estudio estructural del reconocimiento de ARN en Nrd1.** Cristalización de la región de unión a ARN de Nrd1 en estado libre y en complejo con ARN y determinación estructural mediante Cristalografía de rayos X. Caracterización estructural del dominio RRM de Nrd1 y de las regiones flanqueantes. Identificación de los determinantes moleculares de la especificidad de interacción con el ARN.

3. TRABAJOS DE INVESTIGACIÓN

3. Trabajos de investigación

3.1 Trabajos de investigación 1 y 2

3.1.1 Trabajo de investigación 1

Crystallization and Preliminary X-Ray Diffraction Analysis of a Mammal Inositol 1,3,4,5,6-Pentakisphosphate 2-Kinase

Elsa Franco-Echevarría, Julia Sanz-Aparicio, Nathalie Troffer-Charlier, Arnaud Poterszman y Beatriz González

DOI: 10.1007/s10930-017-9717-y

Introducción general e interés de la investigación

La IP₅ 2-K es una enzima de la familia de las IPKs que cataliza la síntesis de IP₆ a partir de IP₅ y ATP. Como se ha comentado en la introducción, esta enzima es muy específica y presenta como característica única dentro del metabolismo de los inositoles: es la única de la familia capaz de fosforilar el anillo de inositol en la posición axial (2-OH). Fruto de su actividad catalítica y el producto de reacción que genera, el IP₆, la IP₅ 2-K está involucrada en multitud de funciones celulares como el desarrollo embrionario y la ubiquitinación de proteínas (ver Introducción, sección 1.3.1.3.1). Por otro lado, en la última década se ha demostrado que la IP₅ 2-K posee una función crucial en la biosíntesis del ARNr, la cual es independiente de su actividad catalítica. Esta función se encuentra mediada por interacciones proteína-proteína (ver Introducción, sección 1.3.1.3.2) (Brehm et al., 2007).

El grupo de investigación en el que se ha desarrollado el presente trabajo determinó en 2010 la estructura de IP₅ 2-K de *Arabidopsis thaliana* (*At*) (Gonzalez et al., 2010) mediante Cristalografía de rayos X. Esta estructura era la única conocida de la familia IP₅ 2-K con anterioridad a este trabajo. Las isoformas de planta y mamífero presentan características bien diferenciadas, con una identidad de secuencia muy baja (24% identidad y 38% similitud). En concreto, las isoformas de ambos reinos presentan inserciones muy diferentes en su secuencia. Además, *At*IP₅ 2-K presenta un sitio de unión a zinc crítico para el correcto plegamiento y la estabilidad de la enzima, el cual no se encuentra conservado en la enzima de mamífero. El estudio de las características específicas de la isoforma de mamífero es primordial debido a la multitud de funciones que posee. La obtención de cristales de proteína es el paso previo clave para la resolución estructural por difracción de rayos X y posterior análisis de la estructura. Para ello, deben cumplirse ciertos requisitos como la obtención de muestras homogéneas, puras y estables que permitan su cristalización. Anteriormente a esta tesis, nuestro grupo de investigación intentó llevar a cabo la expresión de IP₅ 2-K de ratón (*Mus musculus*, *mIP₅ 2-K*) recombinante en bacteria sin éxito. Además de los problemas de expresión que plantea esta familia de enzimas, cabe destacar que también son muy inestables y tienden a precipitar al concentrar, lo que dificulta enormemente la obtención de cantidades suficientes de proteína a una concentración adecuada para su posterior cristalización.

Los objetivos de este trabajo han sido los siguientes:

- Expresión de la proteína completa IP₅ 2-K de *Mus musculus* (*mIP₅ 2-K*) utilizando técnicas recombinantes en bacterias y células de insecto.
- Desarrollo de protocolos de purificación de *mIP₅ 2-K* que permitan obtener muestra de calidad y cantidad suficiente para poder llevar a cabo los experimentos de cristalización posteriores.
- Cristalización de la proteína completa de *mIP₅ 2-K*. Recogida y posterior procesamiento de datos.
- Determinación del estado oligomérico de *mIP₅ 2-K*.

Las conclusiones generales de este trabajo se detallan a continuación:

- La expresión de *mIP₅ 2-K* en bacteria (LSL- *mIP₅ 2-K*) requiere bajas temperaturas y tiempos largos de inducción, mientras que la expresión en células de insecto (GST-*mIP₅ 2-K*) incrementa cinco veces el rendimiento en la purificación.
- La expresión y purificación de *mIP₅ 2-K* permite obtener muestras homogéneas en presencia de IP₆ y en cantidad mínima suficiente para abordar su estudio cristalográfico.
- La proteína completa *mIP₅ 2-K* cristaliza en 0.2M MgCl₂, 0.1 MES pH 6.24 y 12-18% PEG 6K, en el grupo espacial P2₁2₁2₁. El porcentaje de solvente es compatible con la presencia de dos moléculas en la unidad asimétrica.
- Los cristales de *mIP₅ 2-K* expresada en ambos sistemas difractan a una resolución máxima de 4-4.3 Å.
- *mIP₅ 2-K* en ausencia de IP₆ requiere la presencia de sal en el medio. Además, presenta un perfil de degradación muy rápido en las condiciones de cristalización obtenidas.
- Los experimentos de filtración en gel revelan que la enzima es presumiblemente monomérica.

Contribución de la autora de esta Tesis al trabajo:

- Participación en la preparación de los plásmidos adecuados para la expresión de *mIP₅ 2-K* en *E. coli*, y para la preparación de bácmidos en baculovirus.
- Expresión de todas las muestras de *mIP₅ 2-K* producidas en *E. coli*.
- Purificación y cristalización de todas las muestras de *mIP₅ 2-K* usadas en el trabajo.
- Recogida y procesamiento de múltiples conjuntos de datos de difracción en el sincrotrón ALBA, en la línea XALOC.
- Análisis del perfil de degradación de las muestras en las condiciones de cristalización.

A continuación, se muestra en el enlace para descargar el artículo (Trabajo de Investigación 1):

<https://link.springer.com/article/10.1007%2Fs10930-017-9717-y>

3.1.2 Trabajo de investigación 2

The crystal structure of mammalian inositol 1,3,4,5,6-pentakisphosphate 2-kinase reveals a new zinc-binding site and key features for protein function

Elsa Franco-Echevarría, Julia Sanz-Aparicio, Charles A. Brearley, Juana M. González-Rubio y Beatriz González

DOI: 10.1074/jbc.M117.780395

Introducción general e interés de la investigación

La enzima IP₅ 2-K es esencial para un adecuado funcionamiento celular. Posee una importante función catalítica, ya que está involucrada en la síntesis del IP₆, metabolito básico con numerosas funciones en las células e implicaciones en áreas de salud. Además de esta función catalítica, en mamíferos, se ha detectado la participación de IP₅ 2-K en la biogénesis del ARN a través de interacciones con proteínas esenciales (ver Introducción, sección 1.3.1.3.2).

Previamente a este trabajo se tenía conocimiento de la estructura de la IP₅ 2-K de planta (*AtIP₅ 2-K*). No obstante, y dadas las características distintivas que posee la IP₅ 2-K de mamífero en cuanto a su secuencia y funciones reportadas, es imprescindible llegar a comprender las bases moleculares que hay detrás de su biología. Para ello, constituye un paso obligado la determinación de su estructura tridimensional mediante técnicas de Biología Estructural. Estudios previos en nuestro grupo de investigación nos permitieron obtener, tras un gran esfuerzo, cristales de *mIP₅ 2-K* completa (ver Trabajo de Investigación 1, sección 3.1.1) los cuales difractaron a un máximo de resolución de 4.0 Å. Sin embargo, estos datos no permitieron resolver la estructura de *mIP₅ 2-K* usando técnicas de reemplazamiento molecular con la estructura de la IP₅ 2-K homóloga (*AtIP₅ 2-K*). Por lo tanto, a continuación, se realizó un análisis detallado de la secuencia de la familia de IP₅ 2-K con el fin de optimizar la construcción de partida y así obtener cristales con mejor calidad. En este trabajo se ha obtenido la primera estructura de IP₅ 2-K de mamífero en presencia de inositido mediante Cristalografía de rayos X.

Los objetivos que nos planteamos en este trabajo fueron los siguientes:

- Análisis la secuencia de *mIP₅ 2-K* con el objeto de diseñar posibles formas truncadas que permitan obtener cristales con mejor calidad.
- Obtención y cristalización de las muestras truncadas de *mIP₅ 2-K*.
- Recogida de datos de difracción de rayos X y resolución estructural de *mIP₅ 2-K*.
- Análisis y estudio de la estructura de *mIP₅ 2-K* y extracción de conclusiones relevantes.
- Diseño de experimentos bioquímicos que apoyen los resultados estructurales.

Las conclusiones generales de este trabajo se detallan a continuación:

- La forma trucada ΔC - mIP_5 2-K carente de los 21 residuos del extremo C-terminal se expresa 5 veces más que la enzima completa y conserva su actividad.
- ΔC - mIP_5 2-K cristaliza en condiciones similares a la proteína completa y en el grupo espacial $P2_1$. En función del pH (6.25 o 5.50) se obtienen dos formas cristalinas que difractan a una resolución máxima de 3.2 Å y 2.4 Å respectivamente.
- La estructura de mIP_5 2-K se ha resuelto en presencia de ligandos formando un complejo ternario (mIP_5 2-K + IP_5 + ATP) o binario (mIP_5 2-K + IP_6).
- mIP_5 2-K conserva el plegamiento general de las IPKs; presenta dos lóbulos N y C-terminal entre los cuales se une el ATP. El lóbulo C-terminal presenta una región muy extensa formada por hélices α responsable de la unión al IP.
- El reconocimiento del IP en IP_5 2-K de mamífero es similar al descrito previamente en $AtIP_5$ 2-K, presentando diferencias en el reconocimiento de los fosfatos P1 y P3. A su vez se ha identificado un nuevo residuo catalítico, el Asp439.
- Existen tres regiones helicoidales muy diferentes entre las estructuras de mIP_5 2-K de planta y de mamífero, una en el lóbulo N, y dos en el lóbulo C.
- mIP_5 2-K presenta un sitio de unión a zinc nuevo y distinto al de $AtIP_5$ 2-K, formado por 4 cisteínas. El sitio de unión a zinc es esencial para la estabilidad y plegamiento de la proteína.
- mIP_5 2-K presenta una altísima concentración de residuos positivamente cargados en su superficie, por tanto, sitios clave a explorar en posibles interacciones proteína-proteína y/o proteína-ARN.
- La enzima de mamífero es capaz de catalizar la reacción reversa, como se había identificado previamente en la enzima de planta.

Contribución de la autora a este trabajo:

- Diseño de la construcción, expresión y purificación de ΔC - mIP_5 2-K.
- Cristalización de ΔC - mIP_5 2-K y sus complejos.
- Recogida de datos de difracción en estaciones de radiación sincrotrón y posterior resolución estructural.
- Análisis de la estructura de mIP_5 2-K y extracción de las conclusiones más relevantes.
- Diseño de experimentos de mutagénesis a partir de la estructura de mIP_5 2-K.
- Expresión y purificación de todas las muestras nativas y mutadas utilizadas en los ensayos enzimáticos.
- Experimentos de dicroísmo circular de las muestras nativas y mutante C121S.



The crystal structure of mammalian inositol 1,3,4,5,6-pentakisphosphate 2-kinase reveals a new zinc-binding site and key features for protein function

Received for publication, February 7, 2017, and in revised form, April 25, 2017. Published, Papers in Press, April 27, 2017, DOI 10.1074/jbc.M117.780395

Elsa Franco-Echevarría[‡], Julia Sanz-Aparicio[‡], Charles A. Brearley^{§1}, Juana M. González-Rubio[‡], and Beatriz González^{‡2}

From the [‡]Departamento de Cristalografía y Biología Estructural, Instituto de Química-Física "Rocasolano," Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain and the [§]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

Edited by F. Peter Guengerich

Inositol 1,3,4,5,6-pentakisphosphate 2-kinases (IP₅ 2-Ks) are part of a family of enzymes in charge of synthesizing inositol hexakisphosphate (IP₆) in eukaryotic cells. This protein and its product IP₆ present many roles in cells, participating in mRNA export, embryonic development, and apoptosis. We reported previously that the full-length IP₅ 2-K from *Arabidopsis thaliana* is a zinc metallo-enzyme, including two separated lobes (the N- and C-lobes). We have also shown conformational changes in IP₅ 2-K and have identified the residues involved in substrate recognition and catalysis. However, the specific features of mammalian IP₅ 2-Ks remain unknown. To this end, we report here the first structure for a murine IP₅ 2-K in complex with ATP/IP₅ or IP₆. Our structural findings indicated that the general folding in N- and C-lobes is conserved with *A. thaliana* IP₅ 2-K. A helical scaffold in the C-lobe constitutes the inositol phosphate-binding site, which, along with the participation of the N-lobe, endows high specificity to this protein. However, we also noted large structural differences between the orthologues from these two eukaryotic kingdoms. These differences include a novel zinc-binding site and regions unique to the mammalian IP₅ 2-K, as an unexpected basic patch on the protein surface. In conclusion, our findings have uncovered distinct features of a mammalian IP₅ 2-K and set the stage for investigations into protein-protein or protein-RNA interactions important for IP₅ 2-K function and activity.

Inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IP₅ 2-K)³ is a key enzyme of higher inositol phosphate (inositide, IP) metabo-

lism. IP₅ 2-K is present in yeast to mammals and catalyzes the synthesis of phytic acid (inositol hexakisphosphate or IP₆) from IP₅ and ATP (1). A variety of roles have been proposed for this enzyme and its product IP₆ (2) in DNA repair (3), mRNA editing, export and degradation (4, 5), vesicle trafficking (6) and protein ubiquitylation (7). At a molecular level, IP₆ acts as a cofactor for proteins with DNA-dependent protein kinase activity in non-homologous end joining (8) and with GLE1 in mRNA export (4). IP₆ also can act as a folding factor as in the case of adenosine deaminase that participates in editing of mRNA and tRNA (5). In addition, IP₆ is the precursor of inositol pyrophosphates, essential for cellular energy homeostasis, signal transduction control, and apoptosis (9, 10). Mice embryos with IP₅ 2-K deletion do not survive more than a few weeks (11). More recently, a role in ribosomal rRNA synthesis independent of IP₅ 2-K catalytic function has been proposed for the human enzyme (12), which has been shown to colocalize with mRNA either in the nucleus or cytoplasm (13). All these findings increase the potential of this enzyme as an attractive target.

IP₅ 2-K belongs to the inositol polyphosphate (IPK) structural family that include enzymes capable of phosphorylating hydroxyls at different positions of the inositol ring starting from inositol 1,4,5-trisphosphate (IP₃), a well known second messenger responsible for calcium mobilization (14). These phosphorylation events occur in combination with another family of inositol kinases that adopts an "ATP grasp-like" fold (15, 16). Both families together cover a great range of phosphorylation reactions on the six -OH positions of the *myo*-inositol ring and even on those already phosphorylated (15, 17, 18). Many of these enzymes present redundant abilities acting on similar substrates or they bind an inositide in different orientations that is thus phosphorylated in different positions (19). In contrast, IP₅ 2-K is the unique IPK whose physiological role is the phosphorylation of the axial 2-OH position of *myo*-inositol, the other five hydroxyls being in equatorial positions. It is also a

This work was supported in part by Grants BFU2011-24982 and BFU2014-53762-P from the Spanish Ministry of Economy and Competitiveness (to E. F.-E.). The authors declare that they have no conflicts of interest with the contents of this article.

✂ Author's Choice—Final version free via Creative Commons CC-BY license. This article contains supplemental Tables S1–S3 and Figs. S1 and S2. The atomic coordinates and structure factors (codes 5MW8, 5MWL, and 5MWM) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Supported by Biotechnology and Biological Sciences Research Council (United Kingdom) Grants BB/N002024/1 and BB/M022978/1.

² To whom correspondence should be addressed. Tel.: 34-91-5619400; Fax: 34-91-564-24-31; E-mail: xbeatriz@iqfr.csic.es.

³ The abbreviations used are: IP₅ 2-K, inositol 1,3,4,5,6-pentakisphosphate 2-kinase; IP, inositol phosphate; IP₅, inositol 1,3,4,5,6-pentakisphosphate; IP₆, inositol hexakisphosphate; IPK, inositol polyphosphate kinase; LSL,

Laetiporus sulphureus lectin; PEG, polyethylene glycol; TEV, tobacco etch virus; UBF, upstream binding factor; IP₃, inositol 1,4,5-trisphosphate; ICP-OES, inductively coupled plasma-optical emission spectroscopy; mIP, mouse IP; hIP, human IP; r.m.s.d., root mean square deviation; CL, connecting loop; Ins(1,2,4,5,6)P₅, 1,2,4,5,6-inositol pentakisphosphate; Ins(1,3,4,5,6)P₅, inositol 1,3,4,5,6-pentakisphosphate.

very specific enzyme and, together with IP_3 3-K, phosphorylates just one position of inositol (20, 21).

IPK enzymes are classified as a structural subgroup of the protein kinase (PK) family, because they conserve a few features including a fold in two separated lobes (N- and C-lobe) and similar nucleotide recognition mode (22). In addition, the core of the N-lobe and a few involved in catalysis are conserved. In 2004, the first structure of an IPK was described from IP_3 3-K (23). Since then, the structure of at least one member of each IPK subfamily has been reported (23–26). In summary, the IPK family presents a specific and characteristic fold in the C-lobe different from PKs, having a β -sheet core with helical insertions showing a great range of sizes depending on the IPK class. These helical regions are involved in substrate binding and have been named by us as IP-lobe or CIP-lobe (19). Thus, promiscuous IPKs, such as the IP multikinases, have a single helix inserted showing an open active site able to cope with various substrates and products, whereas more specific enzymes as IP_3 3-K and IP_5 2-K show larger helical scaffolds. In fact, IP_5 2-K shows the most elaborated helical region.

The full-length structure of IP_5 2-K from *Arabidopsis thaliana* ($AtIP_5$ 2-K) is the only one known for this subfamily (25). It shows the features described above and shows that it is a zinc metallo-enzyme, for which a structural role has been proposed (25). We subsequently captured different IP_5 2-K conformations by X-ray crystallography (27) displaying open, half-closed, or closed conformations, as the nucleotide and/or the inositol are absent or present in the active site. Extensive work by others and us has identified the residues involved in substrate recognition and catalysis and has depicted the participation of the N-lobe in achievement of a productive conformation (25, 27–30). The structure of $AtIP_5$ 2-K was an important advance for this field, but nevertheless, the specific features of mammal IP_5 2-K remain unknown. The IP_5 2-K family shows moderate sequence conservation across the species. In particular, mammalian enzymes show different insertions and do not conserve the zinc site found in the plant enzyme, making it difficult to obtain a good sequence alignment. From its sequence motifs, putative zinc-binding residues have been proposed in human IP_5 2-K (13). However, if mammal IP_5 2-K is a zinc metallo-enzyme, the putative role for this metal awaits further studies. Because of the significance of IP_5 2-K to proper cell functioning and the many roles of its product IP_6 , it is important to understand the molecular basis that underlies this enzyme function. In this work, we have determined the mouse IP_5 2-K structure in the presence of inositol showing that although this enzyme conserves features with the plant enzyme, it differs significantly in many aspects. Our results define the specific features of mammal IP_5 2-Ks. In addition, we present here valuable information that could help in understanding IP_5 2-K functions beyond its catalytic activity as its role in ribosomal RNA synthesis (12).

Results

Structure of IP_5 2-K mammalian isoform

We have solved the structure of *Mus musculus* IP_5 2-K (mIP_5 2-K) at 2.4 Å resolution (Table 1) from a truncated form lacking the 21 C-terminal residues (ΔC - mIP_5 2-K). Noticeably, mIP_5

2-K crystals were not obtained in the absence of the inositol. The structure for mIP_5 2-K in the presence of one or both ligands, forming binary complexes (IP_6) or ternary complexes ($IP_5 + ATP$), is presented (supplemental Fig. S1). In addition, we present two different crystal forms, including one or two molecules in the asymmetric unit, respectively, the first showing much better resolution (2.4 versus 3.2 Å). As mouse and human IP_5 2-K isoforms share 91% of sequence identity, we propose the structure of the mouse enzyme as a template for the mammalian IP_5 2-Ks.

Mouse IP_5 2-K folds in two lobes, N- and C-terminal lobes, connected by a hinge, thereby conserving the general fold scheme of PKs and IPKs, and in a similar way, both lobes coordinate the nucleotide between them (Fig. 1A). The N-lobe core forms a β -sheet formed by five antiparallel β -strands ($\beta 1$ – $\beta 5$) showing two helical segments. The first helical segment (N-I) harbors $\alpha 1$, equivalent to the helix αC characterized in all protein kinases, whereas the second one (N-II) is a specific insertion different in every IPK subfamily. A role of this region for substrate binding in the IP_5 2-K subfamily has been previously reported by others (29) and by us (25, 27). Regarding the C-lobe, it also presents a β -sheet core formed by five antiparallel β -strands ($\beta 6$ – $\beta 10$). Three helical segments are inserted in the β -sheet core. These segments altogether form a large helical ensemble named the CIP lobe in the structure of $AtIP_5$ 2-K (25), and each of them is consequently named as CIP_I, CIP_{II}, and CIP_{III}. The CIP lobe represents more than half of the protein and is specific to IP_5 2-K enzymes. It creates a scaffold that builds up most of the inositol substrate-binding site. The five loops (CL1–CL5) joining the CIP region to the C-lobe β -sheet core are essential because they play a key role in substrate binding and catalysis (Figs. 1B and 2).

mIP_5 2-K active site and substrate recognition

A general view of IP_5 2-K substrate recognition is shown in Fig. 2A and detailed in supplemental Table S1. The adenine is strongly recognized through polar and hydrophobic interactions with both protein lobes and the hinge connecting them (Fig. 2B). In particular, it forms polar interactions with His-14 and the backbones of Pro-116 and Leu-118. The ribose OHs interact with the C-lobe residues Glu-136 and Arg-209. The triphosphate moiety is tightly bound to the N-lobe of the enzyme through polar interactions and to the C-lobe through two magnesium ions. In particular, phosphate interaction with residue Arg-33, with a flexible loop (G-loop, residues Gly-15–Ser-20) and with an acidic residue (Asp-437) through the magnesium ions, is conserved throughout the PK superfamily and is essential for nucleotide binding and kinase activity.

Regarding the inositol substrate, IP_5 is tightly bound to the enzyme through its five phosphates (Fig. 2C). A total of 14 residues coordinate the phosphate groups, two of them through water molecules and eight of them being lysine or arginine. P1 and P3 are coordinated by both lobes, whereas P4 and P5 are coordinated exclusively by the C-lobe. Residues from the N-lobe involved in P1 and P3 binding come from the segment N-II (Arg-100) and the G-loop (Asn-18 and Lys-19) mentioned above. Residues from the C-lobe involved in coordination of the five phosphates come from the CIP lobe and its CLs. Fig. 2C and supplemental Table S1 show all the polar interactions produced

Structure of a mammal IP₅ 2-K

Table 1

Crystallographic data statistics and refinement

	ΔC -mIP ₅ 2-K	ΔC -mIP ₅ 2-K	ΔC -mIP ₅ 2-K
Ligand modeled	+IP ₅ + ATP	+IP ₆	+IP ₅ + ATP
Crystallization pH	6.25	5.50	5.50
Data collection and processing			
Space group	P2 ₁	P2 ₁	P2 ₁
Unit cell <i>a</i> , <i>b</i> , <i>c</i> (Å)	64.62, 140.76, 68.66	60.55, 71.64, 61.82	60.16, 71.50, 61.20
Unit cell α , β , γ (°)	90.0, 106.5, 90.0	90.0, 111.7, 90.0	90.0, 111.4, 90.0
Temperature (K)	100	100	100
Radiation source	Synchrotron	Synchrotron	Synchrotron
Wavelength (Å)	0.979490	0.979260	0.979260
Resolution range (Å)	49.86–3.20 (3.42–3.20)	44.82–2.54 (2.65–2.54)	71.50–2.40 (2.49–2.40)
No. of observed reflections	134,008 (24,350)	106,382 (13,008)	123,314 (13,608)
No. of unique reflections	19,318 (3482)	16,270 (1970)	18,528 (1961)
Multiplicity	6.9 (7.0)	6.5 (6.6)	6.7 (6.9)
Data completeness (%)	99.3 (99.0)	99.6 (99.7)	97.3 (99.8)
Matthews coefficient (Å ³ Da ^{−1})	2.32	2.32	2.30
No. of molecules in a.u.	2	1	1
Wilson <i>B</i> -factor (Å ²)	83.26	54.65	45.72
Mean <i>I</i> / σ (<i>I</i>)	14.7 (3.3)	12.5 (3.1)	14.9 (3.5)
<i>R</i> _{merge} (%) ^a	10.4 (63.5)	7.6 (58.8)	7.3 (52.8)
<i>R</i> _{pim} (%) ^b	4.3 (25.8)	3.2 (24.6)	3.1 (21.5)
<i>CC</i> _{1/2}	0.99 (0.89)	0.99 (0.99)	0.99 (0.97)
Refinement			
Resolution range (Å)	70.37–3.20	57.45–2.60	56.99–2.40
<i>R</i> _{work} / <i>R</i> _{free} (%)	22.51/24.53	25.34/28.95	24.10/27.12
No. of atoms/Bav (Å ³)	6647/103.43	3348/69.36	3451/52.81
Protein	6575/103.15	3285/69.65	3342/53.178
Ligand	63/153.38	36/64.72	63/44.77
Zinc	4/79.47	1/46.91	1/49.87
Magnesium	2/102.48		2/38.56
Water molecules	3/51.26	26/39.60	39/35.34
Ramachandran plot (%)			
Favored/outliers	88.8/0.0	92.8/0.3	91.7/0.3
r.m.s.d.			
Bonds/angles (Å/°)	0.007/1.18	0.008/1.29	0.006/1.22
Protein Data Bank codes	5MWL	5MWM	5MW8
Missing residues in Protein Data Bank	A: 1–4/99–102/221–225/296–311/ 413–426/465–468 B: 1–7/36–43/97–105/221–226/ 244–251/277–278/298–311/ 412–426/464–468	1–9/40–41/122–128/221–228/ 244–248/298–310/411–419/465–468	1–4/41–42/122–128/221–228/ 244–245/296–310/411–419/465–468

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the measurement of reflection hkl and $\langle I(hkl) \rangle$ is the weighted mean of all measurements.

^b $R_{\text{pim}} = \sum_{hkl} (1/(N-1))^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where N is the redundancy for the hkl reflection.

^c $R_{\text{work}}/R_{\text{free}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$, where F_c is the calculated and F_o is the observed structure factor amplitude of reflection hkl for the working/free (5%) set, respectively.

with the inositol phosphates (P1: Asn-18, Arg-100, Lys-138, Lys-173, and Asn-206; P3: Asn-18, Lys-19, and Lys-441; P4: Gln-449; P5: with Lys-140, Arg-160, and His-164 through water molecules; and P6: Lys-138, Lys-140, Lys-168, Asn-206, and water mediated with Asn-207). The tight and extensive recognition explains the high specificity of this enzyme. Human IP₅ 2-K displays a 0.43 μM K_m value for IP₅ (31) in good agreement with the value obtained for mIP₅ 2-K by us (0.29 μM) (Fig. 3A) and in contrast with the low tens of micromolar values reported for the AtIP₅ 2-K (22 μM) (32). A possible explanation for the different K_m values found in the plant enzyme might be related to some differences found in enzymes from both kingdoms either in the inositol recognition mode or in the constraints introduced by the zinc-binding site (see below).

Six residues from the CLs form the interface of substrate recognition, generating a net of interactions, including the magnesium ions (Fig. 2D). Residues Lys-138 and Asp-400 make direct interaction with P γ and/or the nucleophile 2-OH. The distance between the 2-OH and P γ oxygen is 3.2 Å suggesting an in-line transference mechanism probably in agreement with an associative mechanism. P γ is oriented through a magnesium ion (Mg1) coordinated to Asp-437 and to Asp-439 in a second sphere. Asp-437 also coordinates a second magnesium ion (Mg2) together with Ser-402 through a water molecule.

Comparison of the ternary and binary complexes (IP₅ 2-K/IP₅/ATP versus IP₅ 2-K/IP₆) shows that there is no significant structural variation among them (r.m.s.d. is 0.377 Å for 403 C α atoms). The IP₆ shows similar interactions, with P2 remaining at the substrate interface regions described above. A BLAST search using mIP₅ 2-K sequence and limited to mammals (taxid: 40674) shows that all residues coordinating the nucleotide and inositol are absolutely conserved, with just a couple of residues showing a conservative change in some species.

Unexpectedly, the formation of the ternary complex (mIP₅ 2-K + IP₅ + ATP) was achieved by protein incubation with IP₆ and ADP (supplemental Fig. S1). Therefore, we checked that our crystallized mIP₅ 2-K samples are able to catalyze both the forward (Fig. 3B) and reverse (Fig. 3C) reactions in solution. In agreement, the plant enzyme is highly reversible with an equilibrium constant in the forward “kinase” direction of ~ 14 (20). By ion-pair reverse-phase HPLC, we were also able to confirm the production of ATP from IP₆ and ADP (data not shown).

Mammal IP₅ 2-Ks share a zinc-binding site with a novel structure

mIP₅ 2-K presents two zinc ions in its structure, one in the CIP-lobe (Zn1) and the other close to the hinge region (Zn2) (Figs. 1 and 4). In fact, Zn1 is present in all the crystals obtained,

Structure of a mammal IP_5 2-K

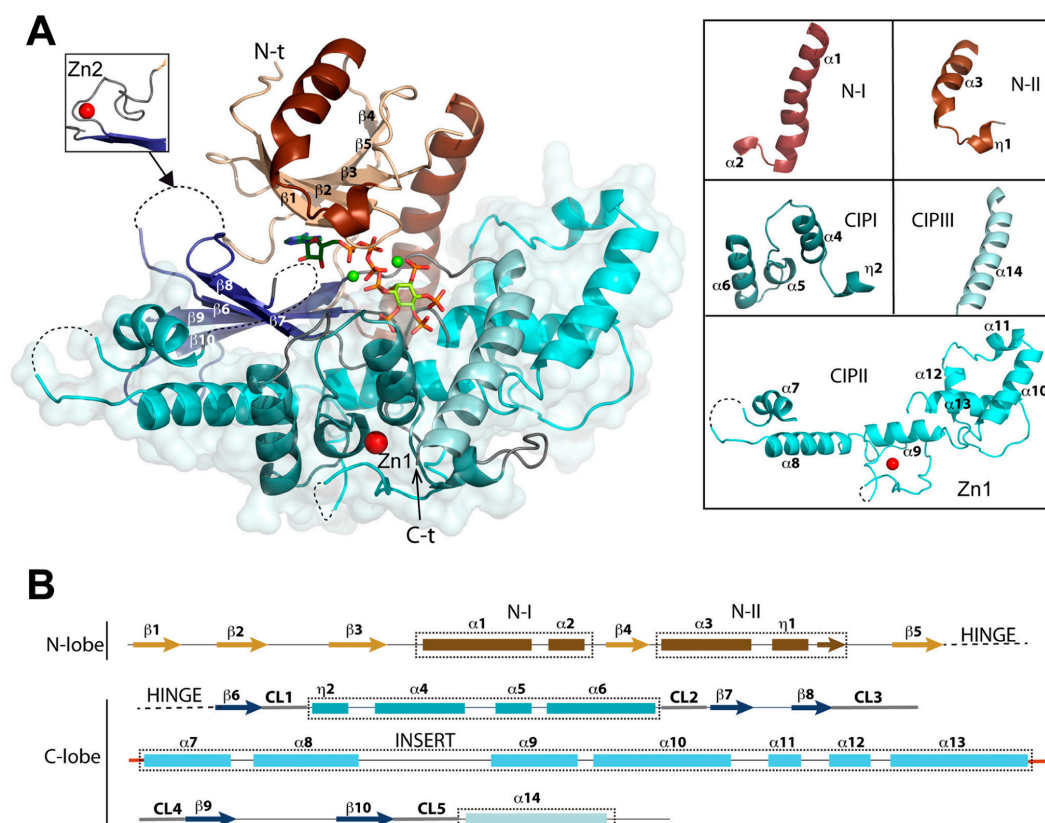


Figure 1. Structure of mIP_5 2-K. **A**, schematic representation of the structure of mIP_5 2-K. The N-lobe and C-lobe are shown in orange-brown and blue colors, respectively. ATP and IP_5 are shown as green sticks, highlighting the oxygen, nitrogen, and phosphorus atoms in red, blue, and orange, respectively. The zinc and magnesium ions are shown as red and green spheres, respectively. The left inset shows a second zinc site found in one of the complexes. The dashed lines show disordered regions. On the right, the five helical segments found in mIP_5 2-K are detailed. **B**, arrows indicate β -strands and rectangles the α -helices. The connecting loops between C-lobe β -sheet and CIP-lobe are highlighted in gray. Conserved sequence motifs within the IPK family are concentrated in the CL β -strands connections (CL1, ¹³⁶EIKPK; CL2, ²⁰⁶QNNRXRF; CL3, variable in sequence and length; CL4, ⁴⁰⁰DCSIM; and CL5, ⁴³⁶LDLDLK).

whereas Zn2 was only detected in the low resolution ΔC - mIP_5 2-K crystals in which the two molecules in the asymmetric unit interact through the hinge region probably fixing a conformation captured by crystallography (supplemental Fig. S2).

The Zn1 site is formed by residues from two CIP lobe elements: Cys-159, Cys-162, and Cys-181 from CIP-I and Cys-291 from CIP-II (Fig. 4A). This site presents the typical zinc geometry and coordination, although it has no homologues in the structural databases using the DALI server (33). The two first cysteines from CIP-I are located in a helix, separated by two residues, and could resemble a partial zinc-finger; however, the fourth ligand breaks any resemblance because it comes from a position very distant in sequence. In fact, Cys-291 comes from a very long loop inserted into two helices that cross over the back of the CIP-lobe (Fig. 5). We have selected Cys-181 and Cys-291 as candidates for the mutagenic study (Fig. 4C). However, there was no expression of the C181S mIP_5 2-K mutant in the conditions reported herein for the wild-type protein. This suggests that the Zn1 site formation could be essential for proper protein

folding. By contrast, C291S mIP_5 2-K mutant expression levels are in the same order as the wild type, being only 2-fold decreased. Despite this, the impact of this mutation on enzyme activity is very high, because the mutated sample retains less than 10% of enzyme activity (Fig. 4C). In this line, it is worth mentioning that the two first cysteines of this zinc site are in helix α_4 , an element that provides four residues for the inositol substrate binding (Fig. 2A). We also consider that this fact could have some effect in obtaining a lower K_m value for the substrate in mIP_5 2-K.

Regarding Zn2, it is created by two residues from the hinge (His-125 and His-129) and one residue from the C-lobe (Cys-410) (Fig. 4B). Mutation of residues His-129 and Cys-410 to serine produces samples with a moderate decrease in enzymatic activity revealing that Zn2 is not critical for protein function (Fig. 4C). A possible explanation for the greater decrease found in the H129S mutant could be its key location in the hinge, an element important for enzyme flexibility.

We have subsequently evaluated the metal content in mIP_5 2-K samples by inductively coupled plasma-optical emission

Structure of a mammal IP_5 2-K

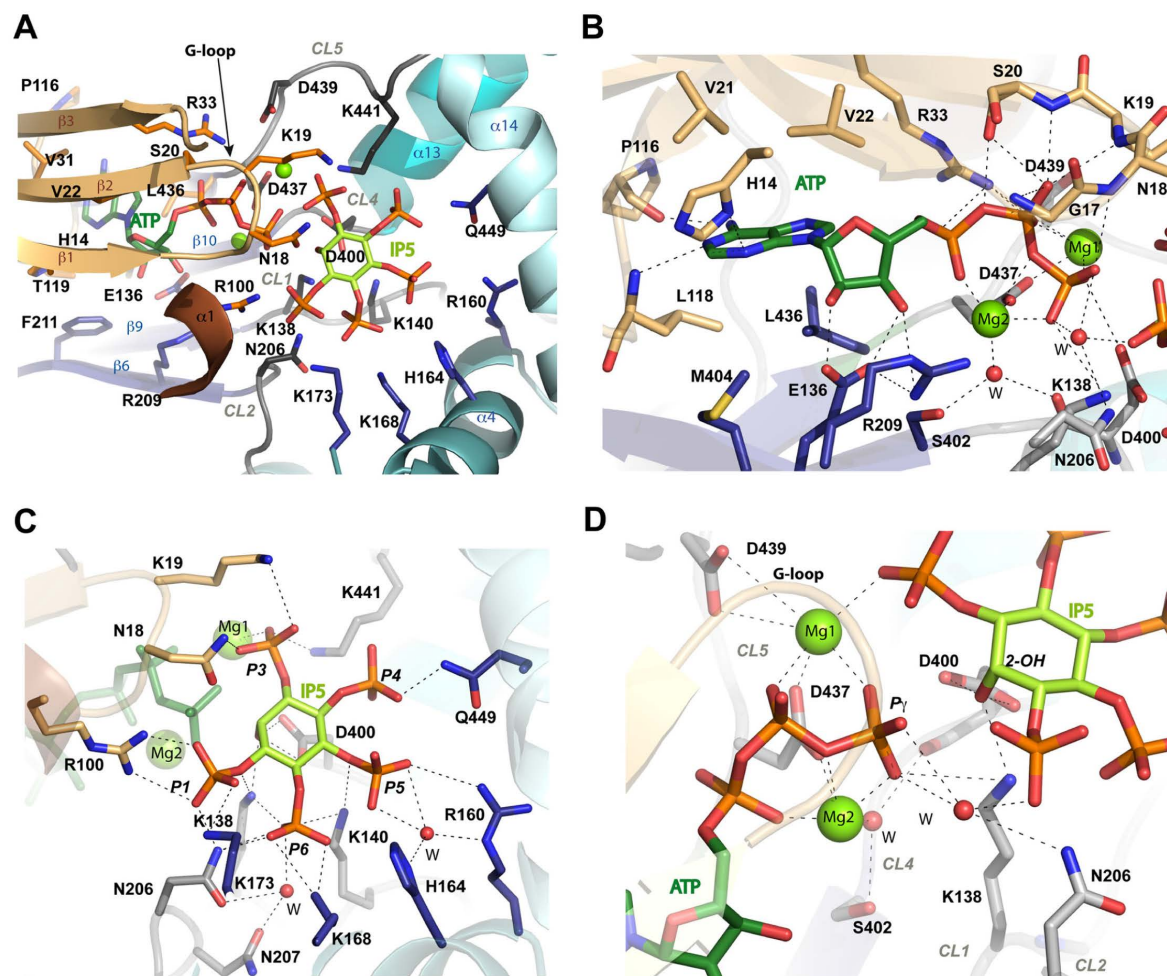


Figure 2. Substrate recognition by mIP_5 2-K. A, zoom of Fig. 1A showing the substrates ATP and IP_5 and the residues involved in their recognition as sticks in light orange (N-lobe), blue (C-lobe), and gray (CLs in C-lobe). B, zoom showing the nucleotide site. Water molecules are shown as red spheres. C, zoom showing the inoside site. D, zoom showing inoside and nucleotide interaction.

spectroscopy (ICP-OES) (Fig. 4D). We confirmed that mIP_5 2-K is a zinc metalloenzyme (supplemental Table S2). Unexpectedly, wild-type samples purified as LSLt-tagged protein exhibit a zinc/protein molar ratio of 0.6:1. An insufficient supply of zinc could also explain the difficulties encountered in the crystallization process of this enzyme due to structural inhomogeneity. However, attempts to add zinc from different salts to the crystallization drops failed because the protein precipitated. We found a zinc/protein molar ratio of 0.8:1, 0.4:1, and 0.7:1 for H129S, C291S, and C410S, mutants, respectively. As expected, mutation on residue Cys-291 (Zn1 site) yields an enzyme with less zinc content than the wild type (Fig. 4D and supplemental Table S2). By contrast, mutation of Zn2-coordinating residues (His-129 and Cys-410) yields samples with no reduction in the zinc content (Fig. 4D and supplemental Table S2). Therefore, we further analyzed the features of the C291S mutant by circular dichroism (CD) (Fig. 4E). We found that both WT (wild type) and C291S mIP_5 2-K samples display a similar far-UV CD

spectra suggesting that they share similar secondary structural elements. However, thermal denaturation followed by CD revealed that the mutated sample exerts an apparent T_m (32 °C) drastically reduced compared with that of the WT sample (43 °C).

In conclusion, our results suggest that Zn1 is necessary for protein folding and stability. Furthermore, the reduced capacity of the C291S mutant to bind zinc correlates with a high reduction in protein activity (Fig. 4, C and D). In agreement, IP_5 2-Ks from mammals show conservation only in the Zn1 site, with its four cysteine ligands being fully conserved. Therefore, Zn1 is key in all mammal IP_5 2-K enzymes, whereas Zn2 could be an artifact of crystallization.

Mammal and plant IP_5 2-Ks show three large structural divergences

Until now, the other IP_5 2-Ks with known structure is that from *A. thaliana* (25). A α superposition of mIP_5 2-K onto *AtIP_5* 2-K (Protein Data Bank code 2xan) overlays 327 resi-

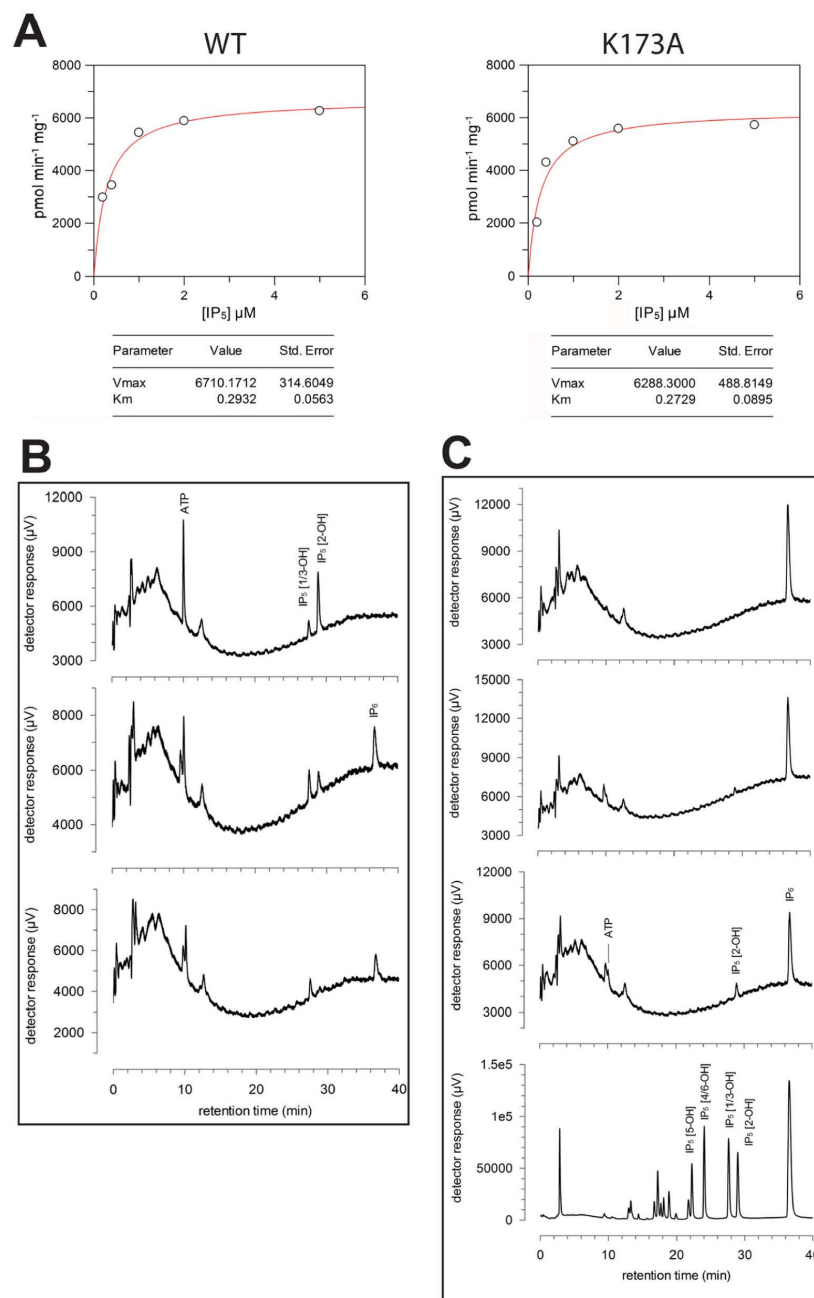


Figure 3. Enzyme assays and reaction catalyzed by mIP₅ 2-K samples. A, kinetic parameters of LSLT-mIP₅ 2-K WT and K173A variant. B, kinase reaction catalyzed by mIP₅ 2-K. Products of enzyme assays resolved by ion-exchange HPLC: substrate, Ins(1,3,4,5,6)P₅ (IP₅) (upper panel); and reaction products after 60 min (middle panel) and 260 min (lower panel) of incubation of IP₅ and ATP with enzyme. C, reversibility of reaction catalyzed by mIP₅ 2-K. Products of enzyme assays resolved by ion-exchange HPLC: substrate, IP₆ (upper panel), reaction products after 110 min (upper middle panel) and 310 min (lower middle panel) of incubation of IP₆ and ADP with enzyme. The lower panel shows IP standards obtained by acid hydrolysis of IP₆.

dues (out of 468 in mIP₅ 2-K) with an r.m.s.d. of 1.2 Å. A good sequence alignment between both enzymes has remained elusive because their sequence homology is not very high (24% identity and 38% similarity), and they present different

insertions. A structural alignment of both IP₅ 2-K isoforms is shown in Fig. 5. Although the topology of the N- and C-lobe cores and substrate-binding region is quite conserved, both enzymes present multiple dissimilarities.

Structure of a mammal IP_5 2-K

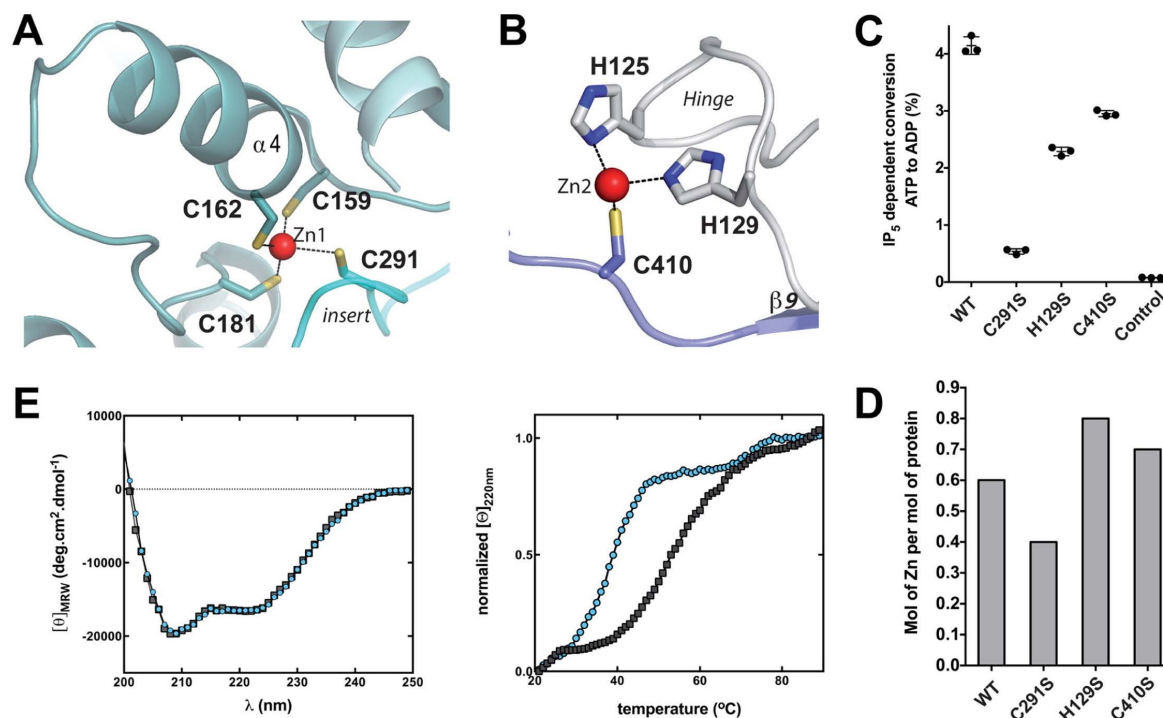


Figure 4. Zinc-binding sites located in mIP_5 2-K complexes. *A*, schematic representation of mIP_5 2-K Zn1 site, and *B*, of Zn2 site. *C*, plot of IP_5 dependent conversion from ATP to ADP of mIP_5 2-K WT and mutants of the zinc-binding residues. Error bars show the standard deviation. *D*, zinc content of mutated mIP_5 2-K samples relative to WT mIP_5 2-K. *E*, far-UV CD spectra (left) and thermal denaturation followed by CD (right) of WT (black squares) and C291S (blue circles) ΔC - mIP_5 2-K.

The most predominant divergences are found within three regions, which show large differences in their sequence, length, and topology (Figs. 5 and 6, *A–D*). The first main difference (D1) is located in N-I region, *AtIP*₅ 2-K having an insertion not present in the mammal enzyme (Fig. 6*B*). A possible function of this segment will be discussed below. The second clear difference (D2) is concentrated within the CIP-II region (Fig. 6*C*). Mouse IP_5 2-K presents a long and flexible loop (Ser-272–Gly-316, 45 residues) that crosses back from the whole CIP (Figs. 5 and 6*C*). Interestingly, this loop is the one that provides a cysteine residue (Cys-291) that completes the Zn1 site present in mIP_5 2-K (Fig. 4). In addition, this loop seems to stabilize the protein, because it packs with several regions of the enzyme, including both lobes, but mainly the CIP lobe. In particular, it makes hydrogen bonds through four residues (Leu-283, Arg-289, Glu-292, and Ser-294) and shows strong hydrophobic interactions through seven residues (Leu-281, Leu-283, Pro-288, Ala-293, Pro-295, Leu-313, and Pro-314) (Fig. 6*E*). These residues are fully or highly conserved along the mammal isoforms. Only residues Leu-283 and Arg-289 show a great variation, and both are involved in hydrogen bonds through their main chain atoms with other parts of the protein (Fig. 6*E*). Finally, a third difference (D3) is shown in a region also located in the CIP-II (Fig. 6*D*). Precisely, *AtIP*₅ 2-K presents a zinc site located in this region (25) that is shorter than and completely different from that found in mIP_5 2-K (Fig. 6, *C* and *D*).

Noticeably, different regions of the two isoforms seem to share roles. On the one hand, D1 and D3, from *AtIP*₅ 2-K and

mIP_5 2-K, respectively (Fig. 6, *B* and *D*), are both stabilizing essential zones of the enzyme, as are $\alpha 1$ (Asn-54–Phe-66, analogous to protein kinases αC) and other CIP residues. Some relevant interactions of these regions are shown in Fig. 6, *F* and *G*. Among them, it is worth mentioning the central role of Tyr-363, Phe-367, and Tyr-368 in mIP_5 2-K (Fig. 6*F*), residues absolutely conserved in the mammal IP_5 2-K isoforms. *AtIP*₅ 2-K presents residues with roles similar to Tyr-363 (Leu-75) and Phe-367 (Trp-69), which interact with equivalent regions of the protein (Fig. 6*G*). We have selected residue Tyr-363 for mutation, because it seems key in the network of interactions as proposed above. Y363A mutation halves enzymatic activity (Fig. 6*H*). Unexpectedly, this decrease in activity is moderate. An explanation is that it corresponds to a single mutation among a multiple net of interactions and in a residue far from the active site. On the other hand, mIP_5 2-K D2 and *AtIP*₅ 2-K D3 (Fig. 6, *C* and *D*) are both involved in generation of the zinc sites, which are not conserved either in sequence, location, or structure between both enzymes. However, both sites seem to have a structural role, although we cannot discard any other additional function. Strikingly, the insertions found in mIP_5 2-K (D2 and D3) interact with each other (Fig. 6*E*). Thus, Tyr-368 in mIP_5 2-K and the following leucine residues (Leu-372 and Leu-374) interact with Leu-281 and Leu-283 located in D2 (Fig. 6*E*). A double mutation in this region (L281A/L283A) reduces slightly the activity in contrast with the >90% decreases caused by the other mentioned mutation on a zinc-binding residue (C291S)

Structure of a mammal IP_5 2-K

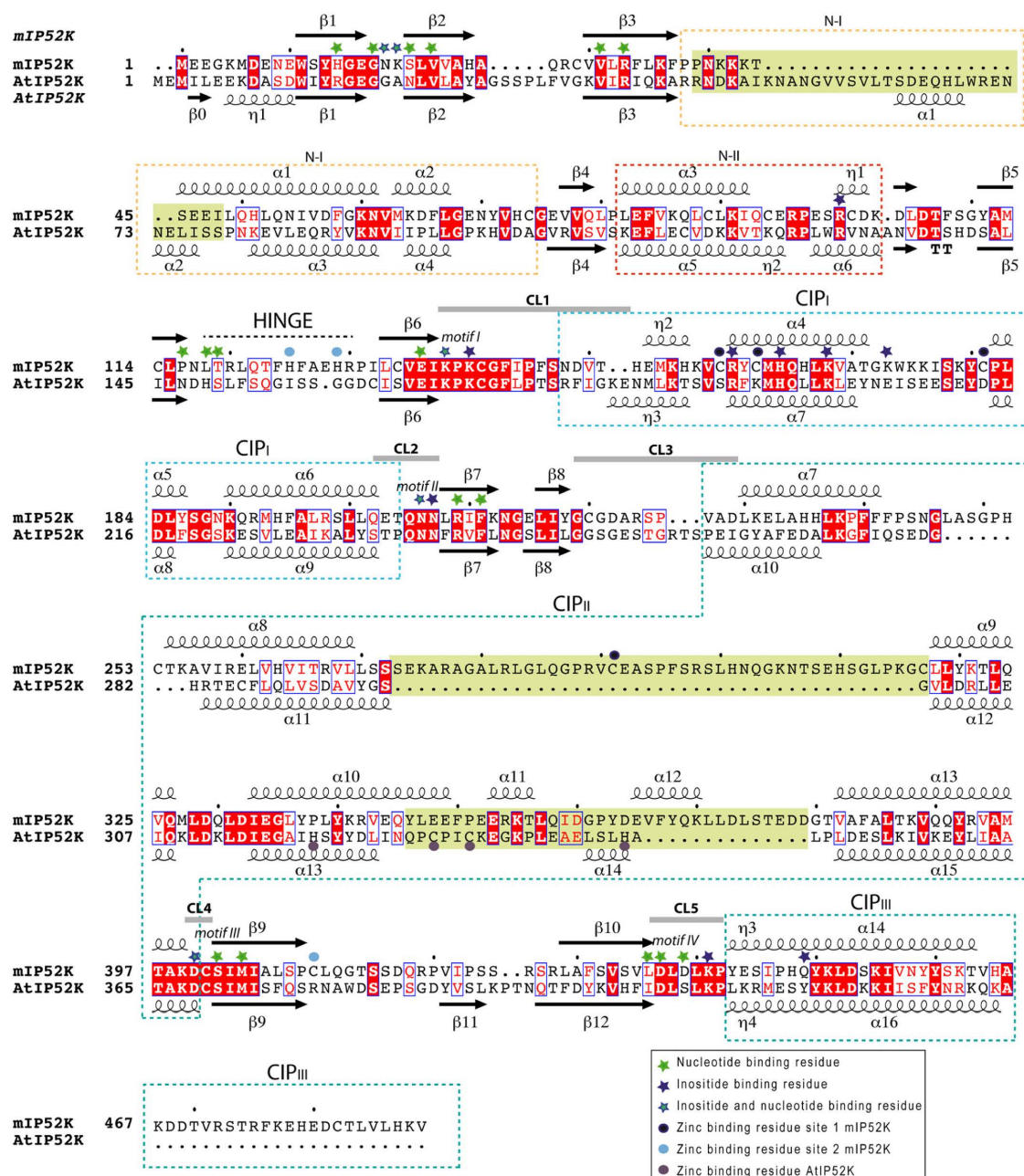


Figure 5. Structural alignment of mIP_5 2-K versus $AtIP_5$ 2-K. The secondary structure is shown at top (mammalian enzyme) and bottom (plant enzyme). Identical regions are shaded red, and similar regions are shown by red letters. Light-green shaded squares show regions with a divergent structure in both enzymes. Helical regions inserted into the β -sheet cores are marked with dashed lined boxes.

within this segment (Fig. 6H). All the mentioned leucine residues show very high conservation, except Leu-283, as mentioned previously. A mutation in this residue would keep the capacity of making hydrogen bonds through its main chain (Fig. 6E) and the slight structural destabilization introduced has no impact in the protein-active site.

Novel findings in mammal IP_5 2-K substrates binding and catalysis

Most residues involved in substrate binding and catalysis are conserved between mammal and plant IP_5 2-Ks (Fig. 5). Nevertheless, we can observe some differences in the inositol P1 and

Structure of a mammal IP_5 2-K

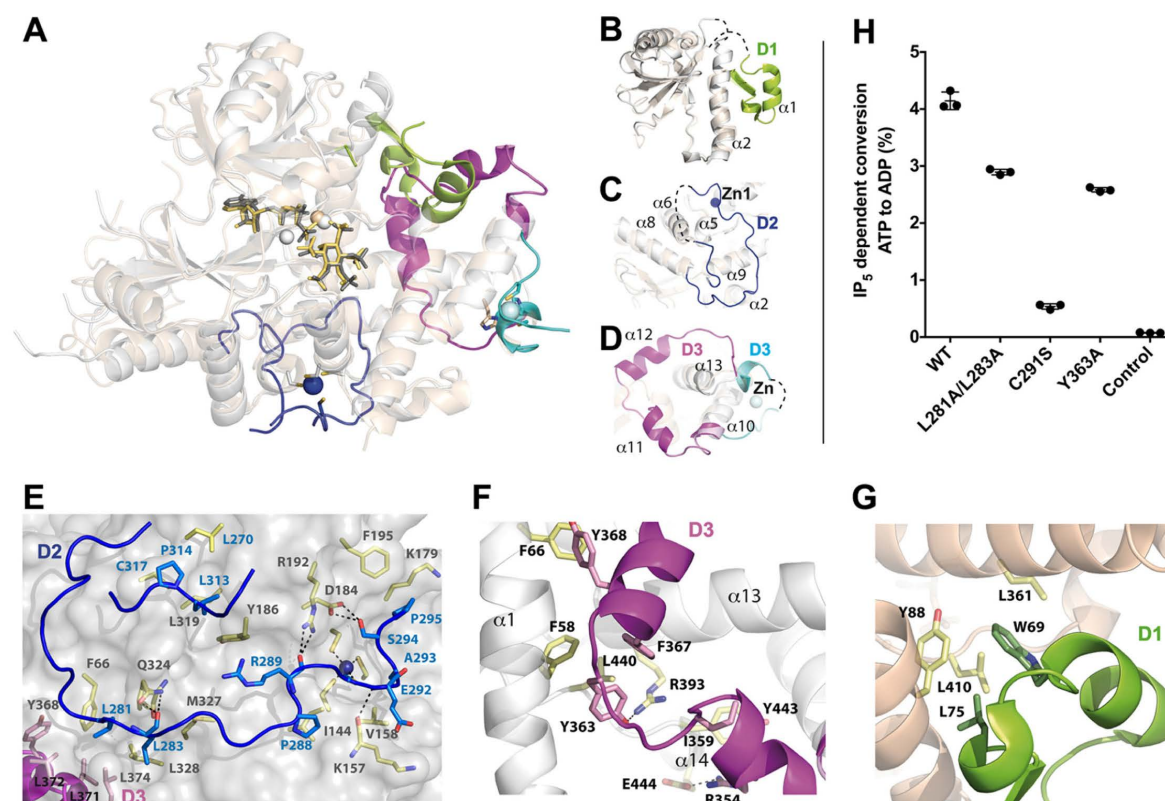


Figure 6. Novel regions of mIP_5 2-K and structural comparison with $AtIP_5$ 2-K. A, schematic representation of mIP_5 2-K (white) and $AtIP_5$ 2-K (wheat) superposed structures. Zinc ions are shown as dark blue (mIP_5 2-K) and cyan ($AtIP_5$ 2-K) spheres. Magnesium ions are shown as spheres in similar color to the isoform to which they belong. The main differences between both samples (D1, D2, and D3) are highlighted in different colors. B, representation of N-lobes from both isoforms highlighting D1 (green) insertion in $AtIP_5$ 2-K. C, piece of CIPII lobe showing D2 in mIP_5 2-K (blue). D, view showing D3 (magenta in mIP_5 2-K and cyan in $AtIP_5$ 2-K). E, intramolecular interactions produced by D2 in mIP_5 2-K. Blue, pink, and yellow sticks show residues from D2, D3, and the rest of the protein, respectively. F, intramolecular interactions produced by D3 in mIP_5 2-K. G, intramolecular interactions produced by D1 in $AtIP_5$ 2-K. Green sticks show residues from D1. H, plot of IP_5 -dependent conversion of ATP to ADP of WT mIP_5 2-K and several mutants of residues located in the different regions. Error bars show the standard deviation.

P3 coordination. As shown, inositide P1 interacts extensively with Arg-100 of mIP_5 2-K (Fig. 2C). The role of the Arg-100 equivalent in $AtIP_5$ 2-K (Arg-130) has been largely argued, and an implication in substrate binding and triggering of a productive protein conformation has been proposed (27, 29, 30). Unfortunately, a construct prepared for the R100A mIP_5 2-K mutant did not show expression in the soluble fraction. However, we can conclude that this residue is structurally similar to Arg-130 in $AtIP_5$ 2-K. In mIP_5 2-K, additional interactions with P1 are produced through the side chain of Lys-173, a residue non-conserved with the plant IP_5 2-Ks but absolutely conserved in mammal enzymes, whereas conservative substitutions can be observed in other vertebrates (Fig. 7). However, mutation of Lys-173 produces an enzyme as active as the wild type (Fig. 7A) showing very similar kinetic parameters (Fig. 3A), suggesting that this interaction is dispensable for substrate binding and it probably might have other implications, as will be commented on later.

Inositide P3 coordinates with the main chain of the G-loop in both enzymes, but different additional interactions are made within the two enzymes. In mIP_5 2-K, P3 also interacts with the

side chains of Asn-18 and Lys-19 G-loop residues, whereas the plant enzyme provides two arginine residues from a different region (Arg-45 and Arg-415) to complete this binding. The double mutation on the G-loop residues Asn-18/Lys-19 (Fig. 7A) has a notable impact in the enzymatic activity, supporting a main role for these G-loop residues absolutely conserved in mammals (Fig. 7C).

We also can observe particular features in catalytic residues in the connecting loops (Fig. 1B). We have prepared mutations on relevant residues from these CLs (K138A, D437A, and D439A) (Fig. 2D), obtaining mutated enzymes with very low activities (Fig. 7B). The role of equivalent residues to Lys-138 and Asp-437 has been largely studied along the PK and IPK families. These residues are responsible for neutralizing the negative charge developed in the transition state and orienting the nucleotide P_γ through magnesium ions, respectively. We observe in mIP_5 2-K that Asp-439 displays a strategic position, helping with magnesium coordination and forming a bridge with Arg-33, a residue involved in the coordination of nucleotide phosphates. Therefore, it seems to provide a proper conformation for essential parts of the enzyme (Fig. 7B). No

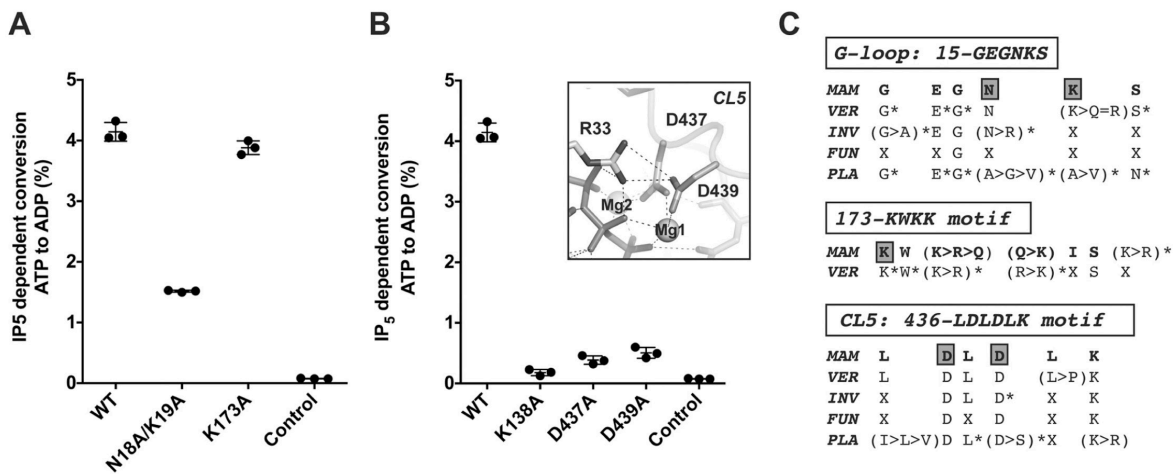


Figure 7. Analysis of relevant residues for mIP_5 2-K function. A, plot of IP_5 -dependent conversion of ATP to ADP by mIP_5 2-K WT and several mutants to determine the effect of selected inositide-binding residues. Error bars show the standard deviation. B, same as A to check the effect of catalytic residues. The inset illustrates the salt bridge between Asp-439 and Arg-33. C, IP_5 2-K sequence conservation across the kingdoms in regions containing the mutated residues (Asn-18 and Lys-19; Arg-173 and Asp-439). Letter X and asterisks denote residues non-conserved or mostly full conserved, respectively. Abbreviations: MAM, mammals; VER, vertebrates other than mammals; INV, invertebrates; FUN, fungi; PLA, plants.

mutagenesis data have been reported on the Asp-439 equivalent residues in other enzymes, because a homologous residue in this position is not present (neither in the IPK or PK families). Asp-439 shows a notable conservation in the whole IP_5 2-K family across the species (Fig. 7C), showing only changes by a serine residue in some plant IP_5 2-Ks, which in turn also coordinates the magnesium atom (25). We show here that the D439A mutation produces nearly inactive enzyme. Finally, the connecting loop CL3 does not show significant conservation between mIP_5 2-K and $AtIP_5$ 2-K, either in length or sequence. This loop is responsible in $AtIP_5$ 2-K for interactions between the N- and C-lobe that might partially regulate the catalysis through opening and closing the active site (27). In addition, this loop makes interactions with two residues directly involved in inositide binding, one of them (Arg-130) shown to be essential for protein activity (27). We think that differences in this loop may also account for the differences observed in the K_m value from each protein. In the mIP_5 2-K structure, this loop is disordered, and therefore interaction with the N-lobe has not been determined (Fig. 1). However, the flexibility found in the mIP_5 2-K CL3 loop is consistent with a dynamic role and the previous proposed functions in catalysis regulation.

Prominent basic patch on mIP_5 2-K surface

The structure of mIP_5 2-K shows a very notable basic patch on its surface, mainly concentrated down the active-site face of the enzyme (Fig. 8A). The enzyme regions that contribute to this patch are $\alpha 6$ and $\alpha 8$ in CIP-I and the large insert found in the CIP-II region. Interestingly, only a few of these basic residues are present or conserved in $AtIP_5$ 2-K (Fig. 8B), which apart from the basic pocket for the inositide binding does not show any significant accumulation of arginine or lysine residues on its surface. By contrast, most of these residues are conserved across the mammalian isoforms. This striking feature could be correlated with mammalian IP_5 2-K localization and/or other

possible functions that this enzyme could present in cells. Particularly outstanding are two basic segments, one formed by residues Lys-175, Lys-176, and Lys-179, and the other containing the residues in the new insert (D2) found in mammals and its preceding helix (Lys-255, Arg-259, Arg-267, Lys-275, Lys-289, and Lys-315). The first segment overlaps in part with the motif $^{173}KW(K/R)(K/Q)$, which is highly conserved in mammals. The second segment encloses one of the main differences (D2) found with respect to plant enzymes. Based on the present finding, we propose that this insert, apart from supporting the zinc binding and stabilizing some protein regions, could be delineating a particular protein surface that could serve as an interface for other partners essential for IP_5 2-K function.

Discussion

Here, we described the first structure of a mammalian IP_5 2-K, a key enzyme in inositol metabolism with multiple impacts in diverse cellular events. Structural knowledge of this enzyme is essential to fully understand its function, although the fact that mammalian IP_5 2-K shows extremely low bacterial expression and a very low tendency to produce suitable crystals has precluded it until now. The structure of mIP_5 2-K presented here shows large structural differences with the $AtIP_5$ 2-K concentrated in the helical regions. Both enzymes are zinc metalloenzymes, the zinc sites showing a different location and structural features. Whereas the plant zinc site is exclusive for its kingdom, the site found in mammals (Zn1) seems to be conserved in all the species except plants. Point mutations of the Zn1 ligands present different effects, going from a null protein expression (Cys-181) to a decreased zinc amount (Cys-291) clearly correlated with a dramatic drop in protein activity and thermal stability. In particular, Cys-291, located in a long insertion (D2), is present in all vertebrates and most invertebrates (data not shown). This insertion is quite unusual because it is unstructured and placed between two contiguous helices (Fig.

Structure of a mammal IP₅ 2-K

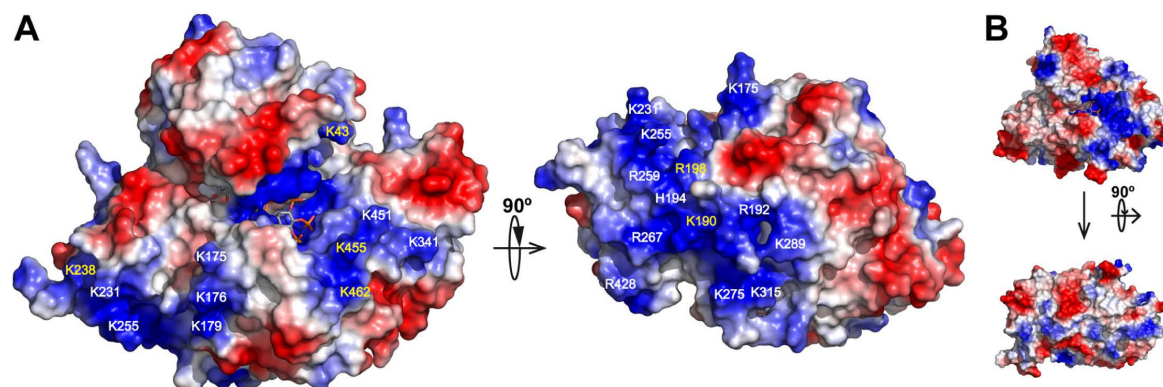


Figure 8. Basic patch on mIP₅ 2-K surface. A, electrostatic surface representation of mIP₅ 2-K showing its exposed basic patch. Residues conserved and non-conserved with the AtIP₅ 2-K are shown in yellow and white, respectively. B, same representation as in A for the AtIP₅ 2-K enzyme.

6C). Its role seems to support the architecture of the CIP lobe to complete the zinc-binding site and to configure a markedly basic protein surface.

Apart from these remarkable differences found, substrate recognition in both isoforms is quite similar. In previous works, AtIP₅ 2-K was shown to display conformational changes upon substrate binding that set the enzyme lobes together producing a closed conformation (27). Studies performed with AtIP₅ 2-K revealed that the IP₅ binding to the C-lobe is stronger, whereas the binding to the N-lobe is necessary for protein activation and conformational change (30). As the Zn1 is located in the CIP-lobe and far from the N-lobe, we suggest that it would affect the preliminary inositide binding events rather than the subsequent dynamic behavior of the protein. The structure solved for mIP₅ 2-K probably corresponds to the close conformation in agreement with the fact that the inositide substrate is present in the structure and the good superposition of mIP₅ 2-K reported here onto the AtIP₅ 2-K closed conformation. We do not know whether the changes reported for AtIP₅ 2-K also occur in the mammal isoforms and whether they are a general behavior of the IP₅ 2-K family. However, the conservation of several elements involved in these changes, as the flexibility of CL3 or inositide binding by N-lobe through Arg-100 or G-loop, suggests that this open-close mechanism could be proposed for mammal IP₅ 2-Ks. In relation to this, we made proteolysis experiments to check whether the substrates protect the enzyme digestion as happened in AtIP₅ 2-K (28). We observed no protection in mIP₅ 2-K, probably due to the fact that its digestion sites are far from regions involved in the conformational change. Moreover, we did not get crystals in the absence of substrates, which could be pointing to some structural changes, although this is not conclusive either.

The structure of AtIP₅ 2-K revealed that IP₅ 2-Ks are the most divergent among the IPK family, because it has the most elaborated CIP-lobe and binds the substrate in a very different orientation to face an axial OH to P γ of phosphate. Similarly, IP₅ 2-Ks are the most divergent enzymes classified inside the PK structural superfamily, which is confirmed in the presented structure. Moreover, we observe that mIP₅ 2-K lacks the N-lobe acidic residue (Glu-91, PKA nomenclature) reported to salt-

bridge to a basic residue (Lys-72, PKA nomenclature) and involved in ATP phosphate coordination. This bridge is a hallmark that identifies the active conformation of protein kinases, and equivalent residues are also present in other IPKs. In contrast, mIP₅ 2-K has a different acidic residue, Asp-439, which stabilizes Arg-33 (equivalent to Lys-72 in PKAs). Therefore, in the IP₅ 2-K family, this bridge could play a similar role to that described in PKs.

Brehm *et al.* (12, 13) have investigated the human (*h*) IP₅ 2-K and, in particular, the possibility of additional functions apart from its catalytic role. They found that *h*IP₅ 2-K colocalizes with mRNA, both in the nuclei and cytoplasm (13). The mutations on a region rich in basic residues, coincident with the ¹⁷³KW(K/R)(K/Q) motif conserved in mammals, show an altered enzyme ability to be exported out of the nuclei. This region is exposed and included within the basic patch of mIP₅ 2-K identified in this work (Figs. 7C and 8). Interestingly, Lys-173 at the beginning of the above motif coordinates the inositide substrate. However, our mutagenesis experiments showed that this residue is dispensable for substrate binding, suggesting that the major role for this motif is played in the translocation process proposed by Brehm *et al.* (13). In addition, *h*IP₅ 2-K was shown to be a structural component of the nucleolus acting as a molecular scaffold in nucleoli and influencing the degree of rRNA synthesis, thereby having a role in rRNA biogenesis. In relation to this, *h*IP₅ 2-K interacts with three proteins (CK2, TCOF, and UBF) that regulate rRNA synthesis. In particular, UBF interacts with a basic region (⁴¹RKK motif equivalent to ⁴¹KKK motif in mIP₅ 2-K) (12) which mutation prevents the UBF translocation out of nucleolus after *h*IP₅ 2-K overexpression. As the authors predict, this region is completely exposed, and we observed that it is in a flexible loop as shown by its poor electron density. We now disclose the high and specific basic region found in the mammalian IP₅ 2-K surface that could be used as a guide to find more target points.

In conclusion, our work provides novel features for the IP₅ 2-K family and its mammalian isoforms. Unexpected protein regions and residues have been identified providing an illuminating picture of these enzymes. The findings comprise the characterization of subtle but important features for substrate

recognition, including unreported catalytic residues for this family, the identification of an unusual and exclusive zinc-binding site, and the conspicuous basic patch on the protein surface. Undoubtedly, the results obtained in this work provide a valuable tool for the design of therapeutics targeted at mammalian IP₅ 2-K with potential implications in health and also to perform IP₅ 2-K functional studies. Beyond the catalytic function, our work also suggests putative regions of interaction of mammalian IP₅ 2-Ks with the cognate partners necessary to accomplish their precise functions.

Experimental procedures

Protein expression and purification

Constructs for full-length IP₅ 2-K recombinant expression either in bacteria (*mipk1*/pKLSLT plasmid) or insect cells were obtained, as described by us (34), from a *mipk1* cDNA (commercial clone bc062167). To produce a truncated *mIP*₅ 2-K enzyme lacking the 21 C-terminal residues (Δ C-*mIP*₅ 2-K), a stop codon was introduced at a position coding for residue 469 of *mIP*₅ 2-K by site-directed mutagenesis and using as template the *mipk1* cDNA inserted into the pKLSLT vector (35). Point and double Δ C-*mIP*₅ 2-K mutants were obtained by site-directed mutagenesis using as template the Δ C-*mipk1*/pKLSLT plasmid. Primers used for construct preparations are shown in supplemental Table S3.

Expression and purification of Δ C-*mIP*₅ 2-K samples fused to LSL— was performed similarly to the full-length samples (34). Briefly, the protein was expressed in *Escherichia coli* BL21 Star (DE3) cells in 2TY medium supplemented with kanamycin (50 μ g ml⁻¹) at 310 K until an A_{600} of 0.9 was reached. Expression was induced with 0.3 mM isopropyl 1-thio- β -D-galactopyranoside for 96 h at 283 K. Pellets were resuspended and sonicated in buffer A (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM DTT) plus 0.2 mM PMSF and 0.05% Triton X-100. The filtrated lysate was diluted 3-fold, loaded onto a heparin column, washed with buffer B (20 mM Tris-HCl, pH 8, 50 mM NaCl, 1 mM DTT), and eluted with a 1 M NaCl gradient. The fusion protein was applied to a Sepharose CL-6B column equilibrated in buffer A and eluted using 200 mM lactose, followed by overnight cleavage with TEV protease (protease/protein mass ratio 1:40) gently rolling at 278 K. Our protein was separated from LSLt and TEV protease by a second heparin column and further purified by size-exclusion chromatography (HiLoad 16/600 Superdex 200 column) equilibrated in buffer A plus 2 mM IP₆, which was included to avoid protein precipitation. All *mIP*₅ 2-K samples used for crystallization were concentrated to around 5–6 mg ml⁻¹ and stored at 193 K. We obtained 1 mg of pure Δ C-*mIP*₅ 2-K per liter of bacteria culture. The purity of all the samples was confirmed by SDS-PAGE. For crystallization and CD analysis, WT and C291S Δ C-IP₅ 2-K samples were purified using this protocol.

Finally, for activity assays, wild-type LSLt- Δ C-*mIP*₅ 2-K and mutants were purified as follows. Clarified and filtrated cell lysate in buffer A was applied to a Sepharose CL-6B column equilibrated in buffer A. After washing with buffer A, the protein was eluted with 200 mM lactose. The sample was diluted 3-fold with 20 mM Tris-HCl, pH 8.0, loaded onto a heparin

column, and washed with buffer B and eluted with a salt gradient. The protein in final buffer C (20 mM Tris-HCl, pH 8.0, 700 mM NaCl, 1 mM DTT) was concentrated to 1–3 mg ml⁻¹ and stored at –80 °C.

Crystallization

All IP₅ 2-K samples used for crystallization were obtained in the presence of 2 mM IP₆. Best crystals obtained for the full-length *mIP*₅ 2-K (expressed either in bacteria or in insect cells) diffracted to 4–4.3 Å (34), and they did not allow the structure solution. Finally, the Δ C-*mIP*₅ 2-K construct allowed us to improve resolution to 3.2 Å from crystals grown in 0.2 M magnesium chloride, 0.1 M MES, pH 6.25, 10% (v/v) PEG 6000 and included 2 mM IP₆ and 2 mM ADP in the protein buffer. A new pH grid screen using the sample in the presence of 2 mM IP₆ allowed us to get better crystals grown in 0.2 M magnesium chloride, 0.1 M sodium acetate, pH 5.5, 16% (v/v) PEG 6000. Soaking experiments in precipitant solutions containing 10 mM IP₆ or 10 mM IP₆/ADP during 3 h yielded the complexes *mIP*₅ 2-K/IP₆ and *mIP*₅ 2-K/IP₅/ATP diffracting to 2.4 Å in the last case. Microseeding technique was necessary to improve the quality of all these crystals. For this purpose, we selected our best crystals and introduced them into 50 μ l of crystallization solution plus a seeding bead. After two cycles of 30 s vortexing and 30 s on ice, we made a seed stock. We streak-seeded the crystallization drops with a whisker using this seed stock. All IP₅ 2-K crystals appeared in a few hours after setting up the crystallization trials, and we observed that the protein is degraded in the crystallization conditions very quickly thus making crystal optimization extremely difficult.

Data collection and structural determination

Crystals were transferred for a few seconds into precipitant solution plus 20% (v/v) glycerol and then flash-cooled in liquid nitrogen. Data from IP₅ 2-K crystals were collected at 100 K in beam line BL13-XALOC of the ALBA Synchrotron (36). Δ C-*mIP*₅ 2-K crystallizes in monoclinic P2₁ space group in two different forms having one (pH 5.5) or two molecules (pH 6.25) in the asymmetric unit (Table 1). Diffraction data were indexed, integrated, and scaled using XDS (37) and merged using Aimless (38) from CCP4 suite (39, 40). Initially, Δ C-*mIP*₅ 2-K monoclinic crystals grown at pH 6.5 (3.2 Å) allowed us to get a partial model using molecular replacement with MOLREP (41) and the structure of *A. thaliana* IP₅ 2-K as a search model (Protein Data Bank code 2XAN). However, the preliminary electron density maps presented many ambiguities, although clearly showed high positive difference peaks for two possible zinc ions. An anomalous map computed with PHENIX (42) showed a strong anomalous signal in those positions (supplemental Fig. S2). Therefore, we tried SAD phasing in combination with MR (MRSAD-Auto-Rickshaw) (43). The heavy atom positions were located using PHASER (44) and refined with MLPHARE (39). The phases obtained were then combined, and density modification was performed with RESOLVE (45, 46) and PIRATE (47). Final electron density maps allowed the building of the whole chain except some exposed loops indicating the flexibility of these regions.

Structure of a mammal IP₅ 2-K

Later, the ΔC -mIP₅ 2-K monoclinic crystals grown at pH 5.5 allowed us to refine the structure of protein complexes with ligands IP₆ and IP₅/ATP to 2.6 and 2.4 Å maximum resolution, respectively. The structures were solved by molecular replacement using MOLREP (41) and the coordinates of ΔC -mIP₅ 2-K described above as a search model. The substrates/products were manually fit into the electron density maps. Although we soaked the crystals with the products IP₆/ADP, the initial electron density maps showed clear density consistent with IP₅ and ATP (supplemental Fig. S1). Then, we checked that our crystallized mIP₅ 2-K samples are able to catalyze both the forward and reverse reactions in solution (Fig. 3, B and C), and, therefore, we modeled the substrates IP₅/ATP in the active site. Model refinement was performed with REFMAC (48) alternating with manual model building using COOT (49). The stereochemistry of the model has been checked with PROCHECK (50). Statistics for all data processing and refinement are summarized in Table 1. Figures of the models were generated with PyMOL (51).

Circular dichroism

CD spectra were recorded using a Jasco-810 spectropolarimeter equipped with a Peltier-thermostatted cell holder. Measurements in the far-UV region (250–200 nm) were performed using the samples WT and C291S ΔC -IP₅ 2-K after protein buffer exchange to 25 mM sodium phosphate, pH 8, and at protein concentrations of 0.1 mg ml⁻¹ (10-mm path length quartz cells; bandwidth, 1 nm; response, 4 s; scan speed, 20 nm min⁻¹). Collected spectra were the average of four accumulations. The data were converted to molar ellipticities after subtraction of the buffer contribution using the average molecular mass per residue (114 Da). Thermal denaturation was monitored by CD measuring the ellipticity changes at 220 nm as the temperature was raised (20–90 °C) at 60 °C h⁻¹. The normalized ellipticity value at each temperature was calculated as $([\theta]_t - [\theta]_{25})/([\theta]_{90} - [\theta]_{25})$, where $[\theta]_t$ is the ellipticity value at temperature t , and $[\theta]_{25}$ and $[\theta]_{90}$ are the ellipticity values at 25 and 90 °C, respectively.

Protein sequence alignments and bioinformatics

IP₅ 2-Ks sequences of all kingdoms have been retrieved using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and Pfam (52) searches. After removing all incomplete sequences or lacking essential hallmarks for kinase function, we had a collection of the following sequences: 102 for mammals, 142 for non-mammal vertebrates, 28 for invertebrates, 158 for fungi, and 203 for plants. Sequence alignments have been performed with the Clustal Omega server (53) and corrected manually with SeaView (54). A structural alignment between AtIP₅ 2-K (code 1xan) and mIP₅ 2-K has been performed with EPSPRIT (55).

Enzyme assays

LSL- ΔC -IP₅ 2-K and ΔC -IP₅ 2-K samples showed comparable activity (data not shown); therefore, we used LSL-tagged samples for analysis of kinetic parameters of WT and mutants. For this purpose, IP₅-dependent conversion of ATP to ADP was determined by HPLC. Assays were performed in 20 mM Hepes,

1 mM MgCl₂, pH 7.3, containing 0.2–5 μ M Ins(1,3,4,5,6)P₅ and 50 μ M ATP in a volume of 50–100 μ l at an enzyme concentration of 4 μ g ml⁻¹. Reactions were stopped by the addition of 50 μ l of 60 mM (NH₄)₂HPO₄, pH 3.8, with H₃PO₄. Aliquots of the reaction products were resolved by ion-pair reverse-phase chromatography (56) with the following modifications: separations were performed on a 100 \times 2.1-mm Agilent X-Bridge C18 (3.5- μ m particle size) column eluted at a flow rate of 0.25 ml min⁻¹. Nucleotide substrates and products (ADP and ATP) were detected at 260 nm, and the extent of conversion of one to the other was determined from the ratio of integrated peaks. The ADP content of the ATP used was less than 0.1% of the ATP peak area. Reaction velocities were calculated assuming 1:1 stoichiometry of consumption of nucleotide and inositol. Experimental data were fitted by non-linear least squares regression to the Michaelis-Menten equation in GraFit (Erithacus Software). Assays were performed in triplicate and the experiment repeated an additional three times with similar results.

Verification of reactions catalyzed by mIP₅ 2-K

We undertook a variety of enzyme assays to determine the identities of products formed by the enzyme. We tested the ability of ΔC -mIP₅ 2-K to catalyze forward “kinase” and reverse reactions.

For the forward kinase reaction, 2.4 μ g of ΔC -mIP₅ 2-K was incubated with 200 μ l of 100 μ M Ins(1,3,4,5,6)P₅ (sodium salt, SiChem, Germany), 40 μ M ATP in 20 mM Hepes, 1 mM MgCl₂, pH 7.3, at 37 °C. At intervals, aliquots were withdrawn and 20 μ l injected onto a 250 \times 3-mm CarboPac PA200 column (Dionex) eluted at a flow rate of 0.4 ml min⁻¹ with a gradient derived from buffer reservoirs containing the following: A, water; B 0.6 M methanesulfonic acid according to the following profile: time (min), % B; 0, 0; 25, 100; 38, 100. The eluate from the column was mixed in a mixing tee with color reagent (0.1% w/v Fe(NO₃)₃·9H₂O in 2% v/v HClO₄) (57) delivered at a flow rate of 0.2 ml min⁻¹ by a second HPLC pump. The combined flow was monitored at 290 nm after passage through a knitted reaction coil.

For the reverse reaction, 2.4 μ g of ΔC -mIP₅ 2-K was incubated with 200 μ l of 100 μ M IP₆ (sodium salt, Merck, Germany), 50 μ M ADP in 20 mM Hepes, 1 mM MgCl₂, pH 7.3, at 37 °C, with subsequent processing as above. The identity of IP₅ and IP₆ products was confirmed by chromatography of standards (SiChem or Merck) and by analysis of an IP₆ hydrolysate obtained by overnight refluxing of IP₆ in 1 M HCl, with subsequent rotary evaporation to remove HCl.

Our Ins(1,3,4,5,6)P₅ (IP₅ 2-OH) substrate contained D- and/or L-Ins(1,2,4,5,6)P₅ (IP₅ 1/3-OH), but our analysis which resolves the two *meso*-compounds (IP₅ 2-OH and IP₅ 5-OH) from the two pairs of enantiomers (IP₅ 1/3-OH and IP₅ 4/6-OH) confirmed that mIP₅ 2-K is an inositol 1,3,4,5,6-pentakisphosphate 2-kinase that does not accept D- and/or L-Ins(1,2,4,5,6)P₅ (mixture unknown) as substrate (Fig. 3B). HPLC traces were exported from Jasco (Great Dunmow, UK) ChromNav software as ascii files and redrawn in GraFit (Erithacus Software).

Inductively coupled plasma-optical emission spectroscopy

Metal analysis of mIP₅ 2-K was performed by optical emission spectroscopy on a Varian Vista Pro ICP-OES. Protein or buffer in which protein was prepared was diluted 125–139-fold in 18.2 megohm·cm water containing 1 N HNO₃. Diluted protein or buffer was subjected to ICP-OES on a machine calibrated with 0–4.0 μM standards of cobalt, copper, nickel, and zinc in 1 N HNO₃. All metals except zinc were close to the limit of detection in the diluted protein sample; they were only slightly above the background in the buffer (supplemental Table S2).

Author contributions—E. F.-E and J. M. G. prepared the constructs for the experiments. E. F.-E expressed, purified, and crystallized all the protein samples. E. F.-E and B. G. solved the protein structures. E. F.-E, J. S.-A., and B. G. analyzed the structural data. C. A. B. performed all enzyme assays and zinc measurements. J. S.-A. participated in the work and results and discussion. B. G. designed the research and wrote the paper. All authors edited the manuscript.

Acknowledgments—Data collection was performed at the XALOC-BL13 beamline at ALBA Synchrotron with the collaboration of ALBA staff. We thank Graham Chilvers, UEA SCI Faculty Analytical Facility, for assistance with ICP-OES. We acknowledge the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INSB-05 and of Instruct, a Landmark ESFRI project.

References

- Stephens, L. R., Hawkins, P. T., Stanley, A. F., Moore, T., Poyner, D. R., Morris, P. J., Hanley, M. R., Kay, R. R., and Irvine, R. F. (1991) Myo-inositol pentakisphosphates. Structure, biological occurrence and phosphorylation to myo-inositol hexakisphosphate. *Biochem. J.* **275**, 485–499
- Shears, S. B. (2001) Assessing the omnipotence of inositol hexakisphosphate. *Cell. Signal.* **13**, 151–158
- Ouyang, Z., Zheng, G., Tomchick, D. R., Luo, X., and Yu, H. (2016) Structural basis and IP6 requirement for Pds5-dependent cohesin dynamics. *Mol. Cell* **62**, 248–259
- Alcázar-Roman, A. R., Tran, E. J., Guo, S., and Wente, S. R. (2006) Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nat. Cell Biol.* **8**, 711–716
- Macbeth, M. R., Schubert, H. L., Vandemark, A. P., Lingam, A. T., Hill, C. P., and Bass, B. L. (2005) Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534–1539
- Boullaran, C., Scott, M. G., Bourougaa, K., Bellal, M., Esteve, E., Thuret, A., Benmerah, A., Tramier, M., Coppey-Moisand, M., Labbé-Jullié, C., Fähræus, R., and Marullo, S. (2007) β-Arrestin 2 oligomerization controls the Mdm2-dependent inhibition of p53. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18061–18066
- Scherer, P. C., Ding, Y., Liu, Z., Xu, J., Mao, H., Barrow, J. C., Wei, N., Zheng, N., Snyder, S. H., and Rao, F. (2016) Inositol hexakisphosphate (IP6) generated by IP5K mediates cullin-COP9 signalosome interactions and CRL function. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3503–3508
- Byrum, J., Jordan, S., Safrany, S. T., and Rodgers, W. (2004) Visualization of inositol phosphate-dependent mobility of Ku: depletion of the DNA-PK cofactor InsP6 inhibits Ku mobility. *Nucleic Acids Res.* **32**, 2776–2784
- Barker, C. J., Illies, C., Gaboardi, G. C., and Berggren, P. O. (2009) Inositol pyrophosphates: structure, enzymology and function. *Cell. Mol. Life Sci.* **66**, 3851–3871
- Chakraborty, A., Koldobskiy, M. A., Sixt, K. M., Juluri, K. R., Mustafa, A. K., Snowman, A. M., van Rossum, D. B., Patterson, R. L., and Snyder, S. H. (2008) HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1134–1139
- Verbsky, J., Lavine, K., and Majerus, P. W. (2005) Disruption of the mouse inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene, associated lethality, and tissue distribution of 2-kinase expression. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8448–8453
- Brehm, M. A., Wundenberg, T., Williams, J., Mayr, G. W., and Shears, S. B. (2013) A non-catalytic role for inositol 1,3,4,5,6-pentakisphosphate 2-kinase in the synthesis of ribosomal RNA. *J. Cell Sci.* **126**, 437–444
- Brehm, M. A., Schenk, T. M., Zhou, X., Fanick, W., Lin, H., Windhorst, S., Nalaskowski, M. M., Kobras, M., Shears, S. B., and Mayr, G. W. (2007) Intracellular localization of human Ins(1,3,4,5,6)P5 2-kinase. *Biochem. J.* **408**, 335–345
- Berridge, M. J., and Irvine, R. F. (1984) Inositol trisphosphate, a novel second messenger in cellular signal transduction. *Nature* **312**, 315–321
- Wang, H., Falck, J. R., Hall, T. M., and Shears, S. B. (2011) Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding. *Nat. Chem. Biol.* **8**, 111–116
- Qian, X., Mitchell, J., Wei, S. J., Williams, J., Petrovich, R. M., and Shears, S. B. (2005) The Ins(1,3,4)P3 5/6-kinase/Ins(3,4,5,6)P4 1-kinase is not a protein kinase. *Biochem. J.* **389**, 389–395
- Onnebo, S. M., and Saiardi, A. (2007) Inositol pyrophosphates get the vip1 treatment. *Cell* **129**, 647–649
- Saiardi, A., Caffrey, J. J., Snyder, S. H., and Shears, S. B. (2000) The inositol hexakisphosphate kinase family. Catalytic flexibility and function in yeast vacuole biogenesis. *J. Biol. Chem.* **275**, 24686–24692
- Shears, S. B. (2004) How versatile are inositol phosphate kinases? *Biochem. J.* **377**, 265–280
- Phillippy, B. Q., Ullah, A. H., and Ehrlich, K. C. (1994) Purification and some properties of inositol 1,3,4,5,6-pentakisphosphate 2-kinase from immature soybean seeds. *J. Biol. Chem.* **269**, 28393–28399
- Irvine, R. F., Letcher, A. J., Heslop, J. P., and Berridge, M. J. (1986) The inositol tris/tetrakisphosphate pathway—demonstration of Ins(1,4,5)P3 3-kinase activity in animal tissues. *Nature* **320**, 631–634
- Cheek, S., Zhang, H., and Grishin, N. V. (2002) Sequence and structure classification of kinases. *J. Mol. Biol.* **320**, 855–881
- González, B., Schell, M. J., Letcher, A. J., Veprintsev, D. B., Irvine, R. F., and Williams, R. L. (2004) Structure of a human inositol 1,4,5-trisphosphate 3-kinase: substrate binding reveals why it is not a phosphoinositide 3-kinase. *Mol. Cell* **15**, 689–701
- Wang, H., DeRose, E. F., London, R. E., and Shears, S. B. (2014) IP6K structure and the molecular determinants of catalytic specificity in an inositol phosphate kinase family. *Nat. Commun.* **5**, 4178
- González, B., Baños-Sanz, J. I., Villate, M., Brearley, C. A., and Sanz-Aparicio, J. (2010) Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositol binding site for axial 2-OH recognition. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9608–9613
- Holmes, W., and Jøgl, G. (2006) Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity. *J. Biol. Chem.* **281**, 38109–38116
- Baños-Sanz, J. I., Sanz-Aparicio, J., Whitfield, H., Hamilton, C., Brearley, C. A., and González, B. (2012) Conformational changes in inositol 1,3,4,5,6-pentakisphosphate 2-kinase upon substrate binding: role of N-terminal lobe and enantiomeric substrate preference. *J. Biol. Chem.* **287**, 29237–29249
- Gosein, V., Leung, T. F., Krajden, O., and Miller, G. J. (2012) Inositol phosphate-induced stabilization of inositol 1,3,4,5,6-pentakisphosphate 2-kinase and its role in substrate specificity. *Protein Sci.* **21**, 737–742
- Gosein, V., and Miller, G. J. (2013) Conformational stability of inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) dictates its substrate selectivity. *J. Biol. Chem.* **288**, 36788–36795
- Gosein, V., and Miller, G. J. (2013) Roles of phosphate recognition in inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) substrate binding and activation. *J. Biol. Chem.* **288**, 26908–26913
- Verbsky, J. W., Wilson, M. P., Kisseleva, M. V., Majerus, P. W., and Wente, S. R. (2002) The synthesis of inositol hexakisphosphate. Characterization of human inositol 1,3,4,5,6-pentakisphosphate 2-kinase. *J. Biol. Chem.* **277**, 31857–31862

Structure of a mammal IP₃ 2-K

32. Sweetman, D., Johnson, S., Caddick, S. E., Hanke, D. E., and Brearley, C. A. (2006) Characterization of an *Arabidopsis* inositol 1,3,4,5,6-pentakisphosphate 2-kinase (AtIPK1). *Biochem. J.* **394**, 95–103
33. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
34. Franco-Echevarria, E., Sanz-Aparicio, J., Troffer-Charlier, N., Poterszman, A., and Gonzalez, B. (2017) Crystallization and preliminary X-ray diffraction analysis of a mammal inositol 1,3,4,5,6-pentakisphosphate 2-kinase. *Protein J.* 10.1007/s10930-017-9717-y
35. Angulo, I., Acebrón, I., de las Rivas, B., Muñoz, R., Rodríguez-Crespo, I., Menéndez, M., García, P., Tatenó, H., Goldstein, I. J., Pérez-Agote, B., and Mancheño, J. M. (2011) High-resolution structural insights on the sugar-recognition and fusion tag properties of a versatile β -trefoil lectin domain from the mushroom *Laetiporus sulphureus*. *Glycobiology* **21**, 1349–1361
36. Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S., and Benach, J. (2014) Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* **21**, 679–689
37. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
38. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292
39. Collaborative Computational Project No. 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
40. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
41. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25
42. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367
43. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1089–1097
44. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
45. Terwilliger, T. C. (1999) Reciprocal-space solvent flattening. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1863–1871
46. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972
47. Cowtan, K. (2000) General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1612–1621
48. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
49. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
50. Laskowski, R. A., M. M. W., Moss DS, Thornton J M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
51. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*. Version 1.6.0.0. DeLano Scientific LLC, San Carlos, CA
52. Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285
53. Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N., and Lopez, R. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43**, W580–W584
54. Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224
55. Gouet, P., Robert, X., and Courcelle, E. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31**, 3320–3323
56. Caddick, S. E., Harrison, C. J., Stavridou, I., Mitchell, J. L., Hemmings, A. M., and Brearley, C. A. (2008) A *Solanum tuberosum* inositol phosphate kinase (StIPK1) displaying inositol phosphate-inositol phosphate and inositol phosphate-ADP phosphotransferase activities. *FEBS Lett.* **582**, 1731–1737
57. Phillippy, B. Q., and Bland, J. M. (1988) Gradient ion chromatography of inositol phosphates. *Anal. Biochem.* **175**, 162–166

3.1.3 Discusión trabajos de investigación 1 y 2

Los trabajos presentados recogen la cristalización y resolución estructural de la proteína IP₅ 2-K de *Mus musculus* mediante Cristalografía de rayos X. Ésta representa la primera estructura conocida de una IP₅ 2-K de mamífero.

En primer lugar, *mIP*₅ 2-K muestra una expresión bacteriana extremadamente baja, así como una tendencia muy reducida a cristalizar, lo que se ha traducido en un proceso de cristalización y resolución estructural arduo y complicado. Los cristales preliminares de *mIP*₅ 2-K presentaron difracción limitada (4 Å en el mejor de los casos) y aparecían en tan solo media hora, tras lo cual se detectó una inesperada digestión de la proteína en las condiciones de cristalización. A pesar de esto, un análisis mediante espectrometría de masas confirmó que la especie cristalizada era la IP₅ 2-K en su versión completa. Curiosamente, las muestras de *mIP*₅ 2-K expresadas en bacterias y en células de insecto cristalizan en el mismo grupo espacial, sin embargo, muestran variaciones en las dimensiones de los ejes b y c de la celdilla, resultando su volumen un 20% menor en los cristales de la muestra expresada en bacteria. Una explicación sería la presencia de modificaciones postraduccionales en la muestra expresada en células de insecto. Esto encaja con los resultados obtenidos en filtración en gel en los que la muestra procedente de células de insecto presenta un volumen de elución menor, si bien en todos los casos este parece compatible con especies monoméricas. Finalmente, con el diseño y preparación de la nueva construcción truncada en el extremo C-terminal (Δ C-*mIP*₅ 2-K) se consiguieron obtener cristales adecuados para lograr con éxito el proceso de resolución estructural. Esto ha sido debido probablemente a dos motivos: por un lado, el nivel de expresión de esta versión mejora sustancialmente lo cual es garantía de un proceso de purificación más eficaz y, por otro lado, la eliminación de 21 residuos en el extremo C-terminal, que se predicen desordenados y no están conservados en otras isoformas, genera una muestra más homogénea.

La proteína *mIP*₅ 2-K ha sido cristalizada siempre en presencia del producto IP₆, dando lugar a complejos binarios (*mIP*₅ 2-K +IP₆) y ternarios (*mIP*₅ 2-K +IP₅ +ATP). Inesperadamente, la formación del complejo ternario que contiene IP₅ se consiguió al incubar la proteína con los productos IP₆ y ADP. Anteriormente ya se había reportado que la enzima de planta es capaz de catalizar la reacción reversa. A la vista de los resultados cristalográficos obtenidos, hemos confirmado mediante ensayos bioquímicos que la enzima de mamífero es también capaz de llevar a cabo la reacción reversa. A partir de la estructura y los estudios de mutagénesis realizados, se ha determinado que la mayoría de los residuos involucrados en la unión al sustrato y la catálisis se encuentran conservados entre *mIP*₅ 2-K y *AtIP*₅ 2-K (ver Figuras 2 y 7 del Trabajo de Investigación 2). Ambos lóbulos, N y C, participan en la unión del inositol, aunque la mayor parte de las interacciones se producen en la región de unión a IP del lóbulo C-terminal. Curiosamente, las diferencias más significativas se observan en la coordinación de los fosfatos P1 y P3 del IP₅, la cual se produce fundamentalmente a través del lóbulo N-terminal (Figura 3.1). En cuanto a los residuos catalíticos, aunque se conservan el par ácido/base descrito (Asp437/Lys138), se observa la intervención de un segundo residuo ácido (Asp439) en la coordinación del magnesio.

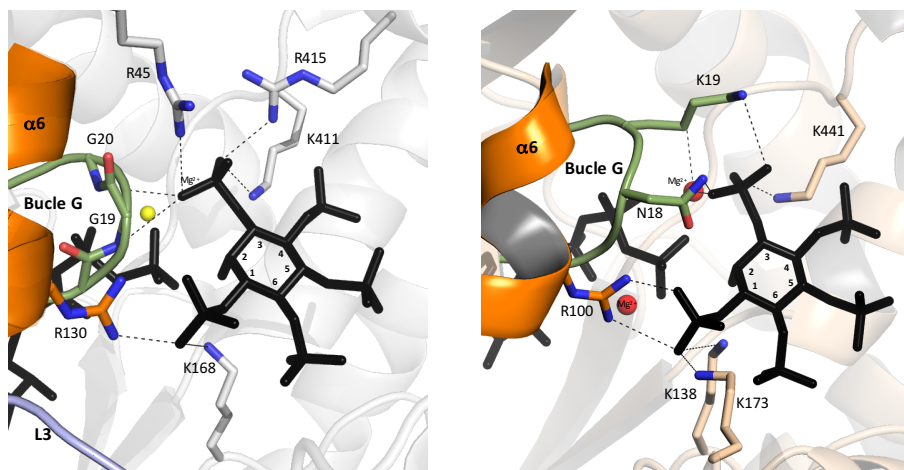


Figura 3.1. Reconocimiento de los fosfatos P1 y P3 del IP en *AtIP*₅ 2-K (izquierda) y *mIP*₅ 2-K (derecha). La estructura general de *AtIP*₅ 2-K se muestra en gris y la de *mIP*₅ 2-K en crema. Los residuos que coordinan el P1 y P3 del IP₅ se representan en bastones, donde pueden apreciarse los elementos específicos de cada enzima. Se destacan en color naranja y verde la hélice $\alpha 6$ y el bucle G, elementos del lóbulo N- implicados en el reconocimiento de P1 y P3 respectivamente.

Como se ha mencionado en la introducción, la enzima *AtIP*₅ 2-K presenta cambios conformacionales muy notables, con una conformación abierta (libre de nucleótido e inositol) con los lóbulos N- y C- separados, una semi-cerrada (unida al nucleótido) y una completamente cerrada (unida al inositol) (ver Figura 1.15 en Introducción). Una superposición de ambas estructuras (*AtIP*₅ 2-K vs *mIP*₅ 2-K) en sus formas unidas a inositol revela que presentan la misma conformación, por tanto, la estructura que se presenta de *mIP*₅ 2-K parece ser equivalente a la forma cerrada. Las tres conformaciones cristalizadas de *AtIP*₅ 2-K revelaron que el sitio de unión a inositol está preformado en la región del lóbulo C pero que no se forma por completo hasta que la enzima adquiere la conformación cerrada tras la entrada del inositol, y el lóbulo N se acerca lo suficiente para unirlo (Banos-Sanz et al., 2012). Estudios bioquímicos posteriores (Gosein et al., 2012; Gosein y Miller, 2013a, b) demostraron que la unión del IP₅ al lóbulo C-terminal es muy fuerte y la unión al N-terminal es necesaria para la activación de la proteína produciéndose el cambio conformacional. Los determinantes de este cambio de conformación son: 1) el bucle G, 2) la hélice $\alpha 6$ y 3) el bucle L3 (Figura 3.2). Los dos primeros son responsables de la coordinación de los fosfatos P1 y P3 del inositol, mientras que el L3 posiciona residuos importantes para la coordinación del inositol en ambos lóbulos. Por tanto, se puede observar cómo los elementos implicados en la unión de IP que más varían entre ambas isoformas son los implicados en el cambio conformacional. En cuanto al bucle L3, este se muestra desordenado en *mIP*₅ 2-K, por lo que no se ha podido establecer una comparación. Cabe destacar que los tres elementos difieren mucho en secuencia entre las dos isoformas. Hasta la fecha, se desconoce si la isoforma de mamífero experimenta cambios conformacionales equivalentes a los descritos en la enzima de planta ya que no se han obtenido cristales en ausencia de sustratos. Sin embargo, la presencia de todos los elementos anteriores sugiere que el mecanismo de apertura-cierre podría extrapolarse también a la isoforma de mamífero. Teniendo en cuenta las diferencias descritas entre estos elementos, probablemente el mecanismo dinámico sea ligeramente distinto.

Por otro lado, ambas isoformas presentan grandes diferencias estructurales en relación a las inserciones que presentan (Figura 3.2). Aunque estas inserciones son distintas, parecen compartir funciones similares como la estabilización de zonas esenciales de la enzima y la generación de los sitios de unión a

zinc. En cuanto al sitio de unión a zinc, este no está conservado en localización, ni en secuencia o estructura. Mientras que el sitio de zinc de *AtIP*₅ 2-K se conserva únicamente en el reino de las plantas, el de *mIP*₅ 2-K parece estar conservado al menos en el reino animal. Sin embargo, en ambas isoformas se ha determinado que el zinc es esencial para el plegamiento y la estabilidad de la proteína. A diferencia de las isoformas de planta y mamífero, que presentan un sitio de unión a zinc propio de cada reino, la estructura de la isoforma de hongo de *IP*₅ 2-K (*CnIP*₅ 2-K) muy recientemente publicada (Oh et al., 2017), carece de un sitio de unión a zinc.

Finalmente, las inserciones particulares encontradas en *mIP*₅ 2-K esbozan una superficie básica muy llamativa que podría servir como una interfaz de interacción con otras proteínas (ver Figura 8 del Trabajo de Investigación 2). La mayoría de estos residuos básicos presentes no están conservados en *AtIP*₅ 2-K, lo que apunta a funciones específicas en la isoforma de mamífero. Como se ha comentado, el grupo del Dr. Shears (Brehm et al., 2007) ha demostrado que la *hIP*₅ 2-K es un componente estructural del nucleólo teniendo un papel claro en la biogénesis del ARN, debido a la interacción con tres proteínas (CK2, TCOF, y UBF) que regulan la síntesis de ARNr. En particular, ya se ha identificado que UBF interacciona con una región básica (motivo 41-KKK de la *mIP*₅ 2-K) (Figura 3.2) completamente expuesta y desordenada en la estructura de *mIP*₅ 2-K. Sería muy interesante conseguir un complejo entre ambas enzimas, así como explorar si el parche básico identificado en este trabajo está implicado en otras interacciones proteína-proteína o proteína-ARN importantes para la localización, función y actividad de *IP*₅ 2-K de mamífero.

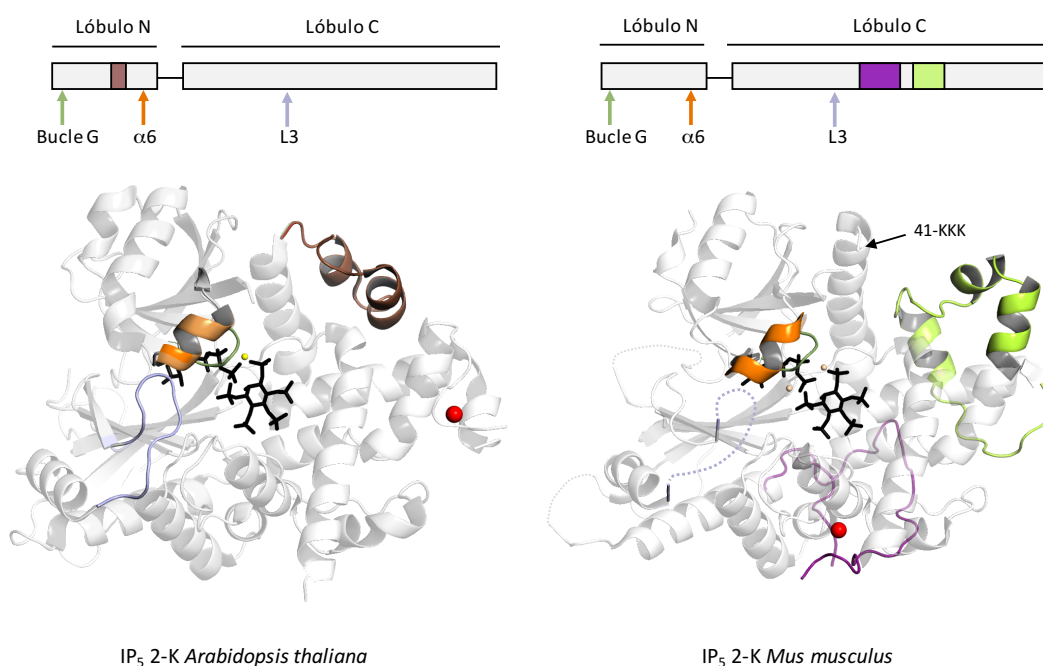


Figura 3.2. Comparación entre las estructuras de *AtIP*₅ 2-K y *mIP*₅ 2-K. Esquema general (superior) y representación de las estructuras (inferior) de ambas isoformas en color blanco. Los elementos responsables del cambio conformacional (bucle G, L3 y hélice $\alpha 6$) se muestran en verde, azul y naranja respectivamente, representándose con flechas en el esquema general superior. La inserción en el lóbulo N-terminal de *Arabidopsis thaliana* (izquierda) se representa en marrón mientras que las inserciones de *Mus musculus* en morado y en verde claro respectivamente, ambas pertenecientes al lóbulo C-terminal. Con una flecha negra se indica la región 41-KKK desordenada en la estructura. El zinc se representa en esfera roja en ambas estructuras.

En conclusión, aunque esta enzima conserva características comunes con la enzima vegetal, difiere significativamente en muchos aspectos y presentan características específicas que se conservan en las isoformas de mamíferos. Sin embargo, se debe seguir investigando para comprender mejor la función de esta enzima por ejemplo mediante la cristalización de complejos proteína-proteína. Por otro lado, el conocimiento estructural de una enzima supone una herramienta muy valiosa para el diseño de inhibidores de su función. En particular, la disponibilidad de moléculas capaces de modificar la actividad de la *mIP₅ 2-K* “*in vivo*” sería también de gran utilidad en el campo de la biología celular y biomedicina.

3.2 Trabajo de investigación 3

The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition

Elsa Franco-Echevarría, Noelia González-Polo, Silvia Zorrilla, Santiago Martínez-Lumbreras, Clara M. Santiveri, Ramón Campos-Olivas, Mar Sánchez, Olga Calvo, Beatriz González and José Manuel Pérez-Cañadillas.

DOI: 10.1093/nar/gkx685

Introducción general e interés de la investigación

En el reino de los hongos se conocen dos tipos de mecanismo de terminación de transcripción, dependiente e independiente de poliadenilación. El primer tipo opera para los transcritos largos (típicamente ARNm), mientras que el segundo ocurre para los transcritos cortos, en general no codificantes (ARNnc). Estos últimos son moléculas funcionales de ARN que, aunque no se traduzcan a proteínas, llevan a cabo actividades fundamentales relacionadas con la regulación de la expresión génica. La terminación de la transcripción de los ARNnc en levaduras es dependiente de un complejo de proteínas de naturaleza modular denominado NNS, formado por tres proteínas: Nrd1, Nab3 y Sen1. Tanto Nrd1 como Nab3 presentan en su secuencia dominios de unión a ARN mientras que Sen1 es una helicasa. Estudios recientes revelan que el complejo Nrd1-Nab3, además de participar en la transcripción del ARN, posee efectos antagónicos a la señalización de RAS, activando la supresión de ciertos genes en condiciones adversas.

Nrd1 en concreto presenta varias regiones con diferentes funciones (ver Introducción, sección 1.5.2). En su parte C-terminal, Nrd1 posee una región de unión a ARN que incluye un dominio RRM implicado en el reconocimiento de los transcritos nacientes que contienen las secuencias terminadoras GUAA/G. En el año 2014 se conoció la estructura de RMN del dominio RRM de Nrd1 (Bacikova et al., 2014), sin embargo, esta ofrece una información imprecisa sobre la región completa de unión a ARN y poco concluyente sobre el mecanismo de reconocimiento del ARN. Es por tanto de gran interés descifrar las bases de la especificidad de Nrd1 por el ARN y así entender mejor su función particular en la terminación de la transcripción del ARNnc. Para ello es necesario acometer el estudio estructural de la región completa de unión a ARN y de sus complejos unidos al mismo.

Los objetivos de este trabajo han sido los siguientes:

- Diseño y producción de construcciones estables de Nrd1 que contienen el dominio de unión a ARN.
- Cristalización del dominio de unión a ARN de Nrd1 en estado libre y en complejo con ARNs específicos de secuencia corta.
- Resolución estructural de los cristales obtenidos mediante difracción de rayos X.

- Resolución estructural del dominio de unión a ARN Nrd1 en solución mediante RMN y comparación con la estructura cristalina.
- Caracterización estructural del dominio completo de unión a ARN de Nrd1: RRM y secuencias contiguas.
- Determinación de las bases de la especificidad y el modo de interacción entre la proteína Nrd1 y el ARN.
- Determinación de las constantes de afinidad complejos Nrd1-ARN y mutante.
- Análisis de la función de los residuos relevantes mediante diferentes técnicas.

Las conclusiones generales de este trabajo se detallan a continuación:

- Las construcciones Nrd1₂₉₀₋₄₆₈ y Nrd1₃₀₁₋₄₈₉ cristalizan en los grupos espaciales P6₅ y P4₃2₁2 y presentan una difracción máxima a 2.3 Å y 1.6 Å respectivamente.
- El dominio de unión a ARN de Nrd1 posee una arquitectura completamente novedosa organizada en dos dominios. Las secuencias contiguas del RRM definen un nuevo dominio denominado SD (“*Split domain*”).
- Las estructuras de Nrd1 en solución y en el cristal son similares, mostrando únicamente cambios significativos en las regiones flexibles N- y C-terminal.
- Las estructuras de los complejos de Nrd1₂₉₀₋₄₆₈ con ARNs de diferente longitud que incluyen el motivo GUAA revelan un reconocimiento específico en el que participan los dos dominios de Nrd1 y que no se extiende más allá del motivo GUAA.
- Nrd1 reconoce de manera específica GUAA ($K_D=2\mu\text{M}$) y de manera semiespecífica GUAG ($K_D= 8 \mu\text{M}$).
- El dominio SD está involucrado en la unión del ARN y es importante para la viabilidad celular.

Contribución de la autora de esta Tesis al trabajo:

- Cristalización de todas las construcciones de Nrd1 presentadas en el trabajo, en su estado libre y en complejo con ARN (GUAA, CGUAAA y UUAGUAAUCC).
- Recogida de datos de difracción de rayos X de todos los cristales descritos en el trabajo.
- Determinación de la estructura tridimensional del dominio de unión a ARN de Nrd1 y de sus complejos con ARN.
- Refinamiento de las estructuras y depositado de datos en el banco de datos PDB.
- Participación en el análisis de las estructuras obtenidas.
- Participación en las medidas de anisotropía de fluorescencia.

The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition

Elsa Franco-Echevarría¹, Noelia González-Polo², Silvia Zorrilla³,
Santiago Martínez-Lumbreras^{4,5}, Clara M. Santiveri⁶, Ramón Campos-Olivas⁶,
Mar Sánchez², Olga Calvo², Beatriz González^{1,*} and José Manuel Pérez-Cañadillas^{5,*}

¹Departament of Crystallography and Structural Biology, Institute of Physical-Chemistry "Rocasolano", CSIC, C/ Serrano 119, 28006 Madrid, Spain, ²Instituto de Biología Funcional y Genómica, IBFG-CSIC, Universidad de Salamanca, ³Department of Cellular and Molecular Biology, Biological Research Center, CSIC, ⁴Department of Chemistry, King's College London, ⁵Department of Biological Physical Chemistry, Institute of Physical-Chemistry "Rocasolano", CSIC, C/ Serrano 119, 28006 Madrid, Spain and ⁶Spectroscopy and Nuclear Magnetic Resonance Unit, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre

Received June 09, 2017; Revised July 18, 2017; Editorial Decision July 23, 2017; Accepted July 25, 2017

ABSTRACT

Transcription termination of non-coding RNAs is regulated in yeast by a complex of three RNA binding proteins: Nrd1, Nab3 and Sen1. Nrd1 is central in this process by interacting with Rbp1 of RNA polymerase II, Trf4 of TRAMP and GUAA/G terminator sequences. We lack structural data for the last of these binding events. We determined the structures of Nrd1 RNA binding domain and its complexes with three GUAA-containing RNAs, characterized RNA binding energetics and tested rationally designed mutants *in vivo*. The Nrd1 structure shows an RRM domain fused with a second α/β domain that we name split domain (SD), because it is formed by two non-consecutive segments at each side of the RRM. The GUAA interacts with both domains and with a pocket of water molecules, trapped between the two stacking adenines and the SD. Comprehensive binding studies demonstrate for the first time that Nrd1 has a slight preference for GUAA over GUAG and genetic and functional studies suggest that Nrd1 RNA binding domain might play further roles in non-coding RNAs transcription termination.

INTRODUCTION

Genetic information is transcribed by three RNA polymerases in eukaryotes, specialised in different types of transcripts. The RNA pol II transcribes the bulk of protein coding RNAs (mRNAs), some 'classical' non-coding genes of well-known function (snoRNAs and snRNAs), and more recently described non-coding transcripts of less known

function (i.e. CUTs, SUTs, etc.) (1). The biosynthesis of all these transcripts is a tightly regulated process that is coordinated with other events of RNA metabolism like nuclear export and quality control/degradation. Transcription termination represents the final step and is performed through two different pathways in budding yeast, dependent on the nature of the transcript (for recent reviews, see (2–6)): the cleavage and polyadenylation pathway, performed by multiprotein factor CPF, processes the mRNAs by cleaving the transcript in the 3'-UTR and adding a poly(A) tail; and the Nrd1–Nab3–Sen1 pathway (NNS) deals with the termination of functional non-coding RNAs (snoRNAs and snRNAs) (7) and promotes the degradation of products arising from Pol II pervasive transcription (e.g. CUTs) (8). This second pathway appears to be less conserved along the eukaryote kingdom than the polyadenylation one and involves three RNA-binding proteins (Nrd1, Nab3 and Sen1) that couple RNA processing and degradation. Nab3 (9) and Nrd1 (10) have been extensively studied and contain RRM domains with sequence specificity for UCUUG (7,11–13) and GUAA/G (10,12,13) respectively. These elements are frequent in snoRNA terminator sequences and their simultaneous recognition is surely boosted by Nrd1–Nab3 heterodimerization (14). In addition, Nrd1 N-terminal CTD interaction domain (CID) interacts specifically with phospho-Ser5 (pS5) repeats of C-terminal domain (CTD) of RNA Pol II Rpb1 subunit (15,16) and with Trf4 from TRAMP complex (17,18) using mutually exclusive interfaces. To further increase this sophisticated landscape of interactions, Nab3 contains low complexity regions and potential oligomerization domains for autoassociation (19).

Currently, there are several structural studies that shed light into this complex network of biomolecular interac-

*To whom correspondence should be addressed. Tel: +34 91 561 94 00; Fax: +34 91 564 24 31; Email: jmperez@iqfr.csic.es
Correspondence may also be addressed to Beatriz González. Tel: +34 91 561 94 00; Fax: +34 91 564 24 31; Email: xbeatriz@iqfr.csic.es

© The Author(s) 2017. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

tions in the NNS complex. The NMR and X-ray structures of Nrd1 CID in complex with pS5 (16) and with Trf4 peptides (18) provide key insights into Nrd1 recruitment to early elongation complexes and the incompatibility between this binding event and TRAMP recognition (thought to occur later). The structure of Nrd1 CID-CTD (pS5) complements those of Pcf11 CID-CTD(pS2) (20) and Rtt103 CID-CTD(pS2) (21) and gives mechanistic clues about how these budding yeast Pol II CTD code readers are recruited to the elongation complex at different stages. Conversely, the protein-RNA recognition in NNS has been less studied at high-resolution. Complexes between Nab3 RRM and UCUU have been solved by X-ray crystallography (22) and NMR (23), but the recognition details of the higher affinity sites (UCUUG) (11) remain unknown. The structure of Nrd1 RNA binding domain has been studied by NMR though only the fold for the RRM counterpart has been determined. (24). In the same work, authors proposed that Nrd1–RNA recognition is semi-specific.

Here, we report the X-ray and NMR structures of Nrd1 RBD. The structures show an unusual fold in which the RRM-flanking sequences define a second well-folded α/β domain (in contrast with previous data (24)). More importantly, we present the X-ray structures of Nrd1 RBD in complex with several RNA sequences containing the GUAA motif, which reveal a unique binding mode that involves both domains. These data, together with our ITC and fluorescence anisotropy binding affinity studies, provide the first explanation at atomic level for Nrd1 specific recognition of GUAA and GUAG previously seen both *in vitro* (10,12,13) and *in vivo* (25,26). We use this detailed structural and biophysical information to rationally design a battery of mutants and study their changes in binding affinity, growth defect phenotypes and snRN13 transcription termination defects *in vivo*. Our work provides high-resolution key structure-function knowledge to progress in our understanding of the mechanism of transcription termination through the NNS pathway.

MATERIALS AND METHODS

Cloning, protein expression and purification

Plasmids used in this work are summarised in Supplementary Table S1. Nrd1 sequences were amplified (KOD polymerase; Novagen) from *Saccharomyces cerevisiae* genomic DNA and cloned in a pET28-modified vector containing an N-terminal fusion cassette (thioredoxin A+6xHis+TEV sequence) ((27) for details). Mutants were obtained with a QuikChange Lightning Kit (Agilent genomics), both in the pET28 and pRS415 plasmid backgrounds using specific DNA oligos (IDT and Macrogen).

Samples of Nrd1 wild type and mutant proteins used in structural and biophysical studies were produced by overexpression in *Escherichia coli* BL21(DE3) cells (Novagen). Cultures were grown at 37°C in LB with 30 $\mu\text{g/l}$ of kanamycin (Sigma-Aldrich). Upon reaching $\text{OD}_{600\text{nm}} = 0.6\text{--}0.8$, cells were transferred to 12°C and induced with 0.5 μM IPTG (Sigma-Aldrich) during 72–96 h. We noticed that expression at higher temperatures result in production of misfolded protein. Uniformly (^{15}N and/or ^{13}C) or selectively unlabeled samples (see specific details in Supplemen-

tary Figure S1) were produced in a similar way in chemically defined media ((27) for details). Selenomethionine substituted Nrd1_{301–489} was expressed similarly, just using minima media supplemented with a mix of inhibition aminoacids and Selenomethionine.

For protein purification, cell pellets were resuspended in buffer A (25 mM Potassium Phosphate pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β -mercaptoethanol and 1 tablet/50 ml of protease inhibitors (Roche)), lysed by sonication and cleared by ultracentrifugation. The supernatant was filtered (0.22 μm) and loaded in a HisTrapTM HP 5 ml column (GE Healthcare), washed with buffer B (as buffer A but with 500 mM NaCl and 30 mM Imidazole) and eluted with buffer C (as buffer A but with 300 mM imidazole). The samples were exchanged to buffer A and cleaved overnight at 16°C with homemade TEV protease (100 $\mu\text{g/ml}$). After complete cleavage, the sample was reloaded on the HisTrap column, further washed with buffer A and selectively eluted with buffer B (TEV protease, cleaved fusion and traces of uncleaved product remain bound to the column). Nrd1 samples were finally polished by gel filtration (Superdex 200 preparative, GE Healthcare) equilibrated in 20 mM Tris–HCl pH 8, 200 mM NaCl and 0.1 mM DTT. Purified Nrd1 samples were concentrated and/or buffer exchanged according to their posterior use.

Protein crystallization

Crystallization experiments with Nrd1_{301–489} and Nrd1_{290–468} constructs were performed at 291K using the sitting-drop vapor-diffusion method and Hampton Research and Qiagen commercial screens using 96-well plates (Innovaplate SD-2 microplates, Innovadyne Technologies Inc). The crystallization conditions were scaled up and refined in 24-well/48-well plates (Hampton Research). Initially, Nrd1_{301–489} (23 mg ml^{-1}) crystallized in 15–18% (v/v) PEG 8000 and 0.1 M Bicine pH 9.3. Selenomethionine (SeMet) modified Nrd1_{301–489} crystals were obtained in very similar conditions consisting of 12% (v/v) PEG 8000 and 0.1 M Bicine pH 9.0. Best condition obtained for a shorter construct, Nrd1_{290–468} (30 mg ml^{-1}), is 0.2 M potassium thiocyanate, 23% PEG 3350 (v/v). A microseeding technique (28) was necessary to improve the quality of these crystals. Apart from these conditions, crystals were obtained in a variety of conditions including a broad range of pHs and using a protein concentration ranging from 20 to 30 mg ml^{-1} . Nrd1 crystals appeared from one to seven days after setting up the crystallization trials.

We tried to prepare crystal complexes between Nrd1 and RNAs with different lengths (4, 5, 6, 10 and 36 bases) by co-crystallization or soaking using all variety of crystals obtained. We had success only with soaking experiments using crystals of Nrd1_{290–468} (24 mg ml^{-1}) grown in 1 M sodium potassium phosphate pH 7.4. The RNA samples were dissolved in the precipitant solution to a final concentration of 4 mM. Nrd1_{290–468} complexes with GUAA, CGUAAA and UUAGUAAUCC RNAs (IBA and IDT) were obtained by slowly adding of 1 μl of the RNA solution to the crystals drops followed by overnight (GUAA) or three hours (CGUAAA and UUAGUAAUCC) incubation.

tion. We only could obtain a single crystal in complex with the longer RNA, since crystals were quickly crashed after its addition. For complex with CGUAAA, Nrd1_{290–468} was dialyzed in 20 mM sodium potassium phosphate pH 8, 200 mM NaCl, 1 mM DTT buffer to remove the tris molecule found in the structure.

Data collection and processing

Prior to data collection, all crystals were transferred for a few seconds to the crystallization solution plus 20–25% (v/v) ethylene glycol or 25% (v/v) glucose and then flash cooled in liquid nitrogen. Diffraction data for Nrd1_{301–489} crystals were collected at the European Synchrotron Radiation Facility (ESRF) (beamline ID23-1) (Grenoble, France), while data sets for the SeMet Nrd1_{301–489} and Nrd1_{290–468} (apo and soaked with RNA) were collected in ALBA (beamline BL13-XALOC) synchrotron facilities at -173°C (Supplementary Table S2). Data processing was performed with XDS package (29) and merging with Aimless (30) from CCP4 suite (Collaborative Computational Project, Number 4, 1994). All constructs used crystallized in the tetragonal space group $P4_32_12$, whereas the shorter construct Nrd1_{290–468} also crystallized in the hexagonal group $P6_5$, in all cases with one molecule in the asymmetric unit (Supplementary Table S2). The structure of Nrd1_{301–489} was solved by single-wavelength anomalous dispersion (SAD) technique from the SeMet-Nrd1_{301–489} dataset collected at the Se fluorescence peak wavelength. Two of three expected selenium positions were located using SHELX (31). Buccaneer was used for preliminary model building (32). Subsequently the structure of non-substituted Nrd1_{301–489} was obtained at 2.3 Å resolution by difference Fourier synthesis using the experimental model above. The shorter construct Nrd1_{290–468} hexagonal crystals diffracted to 1.6 Å resolution and its structure was obtained by molecular replacement with MOLREP (33) using Nrd1_{301–489} refined model as a template. Finally, the Nrd1_{290–468} tetragonal crystals allowed us to get three different RNA-Nrd1 complexes (Supplementary Table S2), which diffracted to 2.45 Å maximum resolution in the best case (GUAA complex). The structure was solved by molecular replacement using the coordinates of Nrd1_{290–468} as a search model. The RNAs were manually built into the electron density maps using COOT (34). Model refinement in all cases was performed by alternating cycles of automatic refinement with REFMAC (35) and manual building with COOT. Statistics for all data processing and refinement are summarized in Supplementary Table S2. The electron density maps allowed building of all chain except for some N- and C-terminal residues (Supplementary Table S2). The stereochemistry has been checked with PROCHECK (36). The figures of the models were generated with PYMOL (37).

NMR

Nrd1 samples (100–800 μM) were prepared in NMR buffer (25 mM Potassium Phosphate pH 6.5, 25 mM NaCl, 1 mM DTT and 10% D_2O) and experiments acquired at 25°C on cryoprobe-equipped Bruker AV800 MHz spectrometer. Resonance assignments (^1H , ^{15}N and ^{13}C) were obtained

with a battery of triple and double resonance 3D experiments (HNCA, HNCO, CBCA(CO)NH, HCCH-TOCSY) (38,39) that were processed with NMRPipe (40) and analysed with CcpNmr Analysis (41).

The Nrd1_{290–468} NMR structure was calculated from NOE-derived distance restraints (Supplementary Tables S3, S4 and Supplementary Figure S2) obtained from 2D NOESY and 3D $^{13}\text{C}/^{15}\text{N}$ -HSQC-NOESY experiments and from a set of 2D ^{13}C - ^{15}N (F1-filtered) NOESY and 2D ^{13}C - ^{15}N (F1/F2-doublefiltered) (42) NOESY experiments recorded on samples with amino acid selective reverse unlabeled (Supplementary Figures S1 and S2C). Backbone dihedral angle restraints (ϕ and ψ) were obtained with TALOS+ (43) from $\text{C}\alpha$, $\text{C}\beta$, CO and NH chemical shifts. Structures were calculated with CYANA 2.1 (44), starting from 50 randomly-generated conformers, using a restrained simulated annealing protocol. The final 20 conformers with lower target function and having no distance (>0.2 Å) and angle ($>5^{\circ}$) violations were selected and subjected to energy minimization with AMBER.

Nrd1_{290–468}-GUAA complex formation was followed by ^1H - ^{15}N -HSQCs by progressively adding the RNA (IBA), in the same NMR buffer as free protein, until reaching slight excess of (1:1.2). A series of 3D HNCO, 3D HNCA and 3D CBCA(CO)HN were taken to assign Nrd1 bound state. The chemical shift perturbations (CPS) were calculated for each NH peak with the equation: $\Delta\delta^{\text{av}} = (1/2 \cdot ((\Delta\delta_{\text{H}})^2 + (0.2 \cdot \Delta\delta_{\text{N}})^2))^{1/2}$.

Fluorescence anisotropy

Measurements were performed in a BMG Polarstar Galaxy plate reader essentially as described in (27). The temperature was 26°C , the concentration of the fluorescein labeled oligonucleotides (IDT) was 40 nM, and the buffer was 20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT pH 8.0. A 1:1 binding model compatible with the experimental data was fitted to the isotherms using BIOEQS software as previously described (27). Errors in the fitting parameters were obtained by confidence limit testing, using the same software, at the 67% confidence level. The intensity of the emission of the fluorescein dye remained essentially unchanged in the presence of the protein.

Isothermal titration calorimetry

ITC experiments were performed at 25°C using a MicroCal iTC200 (Malvern Instruments, UK) calorimeter. Protein and RNA samples were prepared in 20 mM Potassium Phosphate buffer pH 7.0, 150 mM NaCl and 2 mM β -mercaptoethanol. CCGUAACC (230 μM) and CCGUAGCC (470 μM) RNAs (IDT) were titrated into Nrd1_{290–468} at 24–25 and 46–48 μM concentration, respectively, placed in the 200 μl sample cell. The reference cell was filled with distilled water. Titration experiments consisted of 19 injections of 2 μl (with a first injection of 0.4 μl) separated by 150 s to allow thermal power to return to baseline. For homogeneous mixing in the cell, the stirring speed was 1000 rpm. Data were analysed with Origin 7.0 (OriginLab) using a one-site binding model. The experiments were carried out in duplicate. Control experiments of dilution of the

RNAs into buffer were performed and the resulting heats were subtracted from the protein-RNA titration data.

Yeast strains

Plasmids and strains used are listed in Supplementary Table S1. Strains construction and other genetic manipulations were performed following standard procedures (45). Briefly, we used the EJS101-9d strain (10) to construct all *nrd1* mutants analysed in this study. In this strain, *NRD1* has been replaced by the *HIS3* marker, and a genomic wild-type copy of *NRD1* is expressed from a centromeric *URA* plasmid (pRS316). We transformed EJS101-9d strain with the centromeric *LEU* plasmid pRS415, containing either wild-type (pRS415-*NRD1*) (15) or *nrd1* mutant genes. Then, the transformant were selected in -URA-LEU selective media, and thereafter grown in 5-FOA containing media, forcing pRS316-*NRD1* loss, and expression of *NRD1* wt and *nrd1* mutant versions (obtained as described above) from the *LEU* plasmid.

In the case of *nrd1* mutations that caused lethality, we carried out another approach to confirm it. We used then the DLY883 strain (8), where *NRD1* is expressed under the control of a GAL promoter. Thus, in glucose containing media, *NRD1* expression is abolished. We transformed DLY883 with an empty *LEU* plasmid (pRS415) or with the above mentioned plasmids, bearing *NRD1* wt or *nrd1* gene mutants. Yeast transformants were selected in 2% GAL-LEU media and then transferred to a 4% glucose-LEU media to turn off Nrd1 expression.

The EJS101-9d and DLY883 strains, as well as the pRS415-*NRD1* plasmid used as template to generate all *nrd1* mutated genes, were kindly provided by S. Buratowski.

Northern blot analysis

Total RNA was isolated using Qiagen Purification Kit from cells grown at 28°C and after a shift to 37°C for 1 h. Approximately 25 µg of total RNA was loaded onto a 1.5% MOPS-formaldehyde agarose gel, and then RNA was transferred onto nylon membranes. Prehybridization and hybridization were performed in a solution containing 1% SDS, 1 M NaCl and 10% dextran sulphate at 60°C. In the case of the prehybridization, it was carried out for at least 2 h, and in the case of the hybridization, the solution contained in addition the radiolabeled probe (*SNR13* or *ADH1*) and it was carried out overnight. Thereafter, the membrane was twice washed with 2× SSC, 1% SDS at 60°C for 5 and 30 min, respectively and once with 0.1× SSC for 1 h at room temperature. After washes, the membranes were exposed to X-ray film.

RESULTS

Nrd1 RBD forms a novel structure with two domains

The sequence alignment along Nrd1 orthologues (Supplementary Figure S3) shows high conservation in regions flanking the canonical RRM domain (339–410). The ¹H-¹⁵N HSQC spectra of Nrd1_{301–489} and Nrd1_{290–468} constructs are very similar, with differences in the N- and C-terminal regions (Supplementary Figure S4). The ¹³C conformational shifts (Supplementary Figure S5) revealed regular secondary structure elements (β-sheets and α-helices)

outside the RRM. In the Nrd1_{301–489} the C-terminal segment 468–489 is disordered (random coil ¹³C chemical shifts and sharp linewidths; Supplementary Figures S4 and S5). Given these observations we decided to setup crystallization experiments with constructs including and lacking this fragment.

Nrd1_{301–489} and Nrd1_{290–468} crystallized under different conditions and two different space groups, *P*₄₃₂₁2 and *P*₆₅ (Supplementary Table S2). The structures of the two constructs are nearly identical with remarkable differences at the N-terminus (before Asp308) and the C-terminus (after Ile463). The Nrd1 RBD adopts a dumbbell-like structure with two α/β domains occupying the positions of the weights (Figure 1A). The RRM domain (residues 339 to 407) shows the classic α/β topology (Figure 1B). The second domain is built up from two discontinuous segments (304–338, 408–464) that flank the RRM (Figure 1A&B). Topologically, its structure is accurately described as an insertion of the RRM into one of the loops of this second domain. For simplicity, we will refer to this second domain as ‘split domain’ (SD). The SD architecture is different to that of the RRM. It contains a mixed 5-stranded β-sheet with the two central strands in parallel orientation, and an α-helix that runs nearly perpendicular to the β-sheet (Figure 1B). Among the RRM structures having N- or C-terminal extensions, the structure of the Nrd1 RBD represent an utterly new configuration. While we were preparing this manuscript the structure of Seb1 RBD (Nrd1 homolog in *S. pombe*) was published (46) showing an equivalent fold to our Nrd1 structure.

The crystal structure of the Nrd1_{301–489} construct displays a longer C-terminus (464–471) than the structures of the Nrd1_{290–468} construct (Supplementary Figure S6). These residues fold back and loosely interact with the SD, in agreement with their high temperature factors and sharp NMR signals. The two crystal structures have different conformation at the N-terminus. In the Nrd1_{290–468} construct Leu305 interacts with Trp353 and the polypeptide chain can be traced up to Asp302 (Supplementary Figure S6A) whereas for the construct Nrd1_{301–489} the electron density is lost before His304 and Trp353 shows a flipped out conformation (Supplementary Figure S6B).

We also obtained the NMR structure of Nrd1_{290–468} in solution (Figure 1C and Supplementary Table S3) that is essentially equivalent to the crystal one (Supplementary Figure S7A) but reveals great differences from the previously published NMR structure (24), which shows an RRM with very similar fold and an SD with no tertiary fold (Supplementary Figure S7B). We think that this discrepancy probably responds to differences in constructs used or in protein expression protocols. The conformation of the N-terminus (Supplementary Figure S6C) shows the Leu305/Trp353 interaction seen in the crystal structure of the equivalent construct (Supplementary Figure S6A), and an additional contact with Phe298 (disordered in the crystal). The Trp353 Nε1–Hε1 signal in the ¹H-¹⁵N HSQC spectrum of Nrd1_{290–468} is duplicated, a minor peak of 20% population, evidencing conformational heterogeneity. The position of this minor crosspeak coincides with the Trp353 Nε1–Hε1 peak in the Nrd1_{301–489} spectrum (Supplementary Figure S4). Since this construct lacks Phe298, we interpret

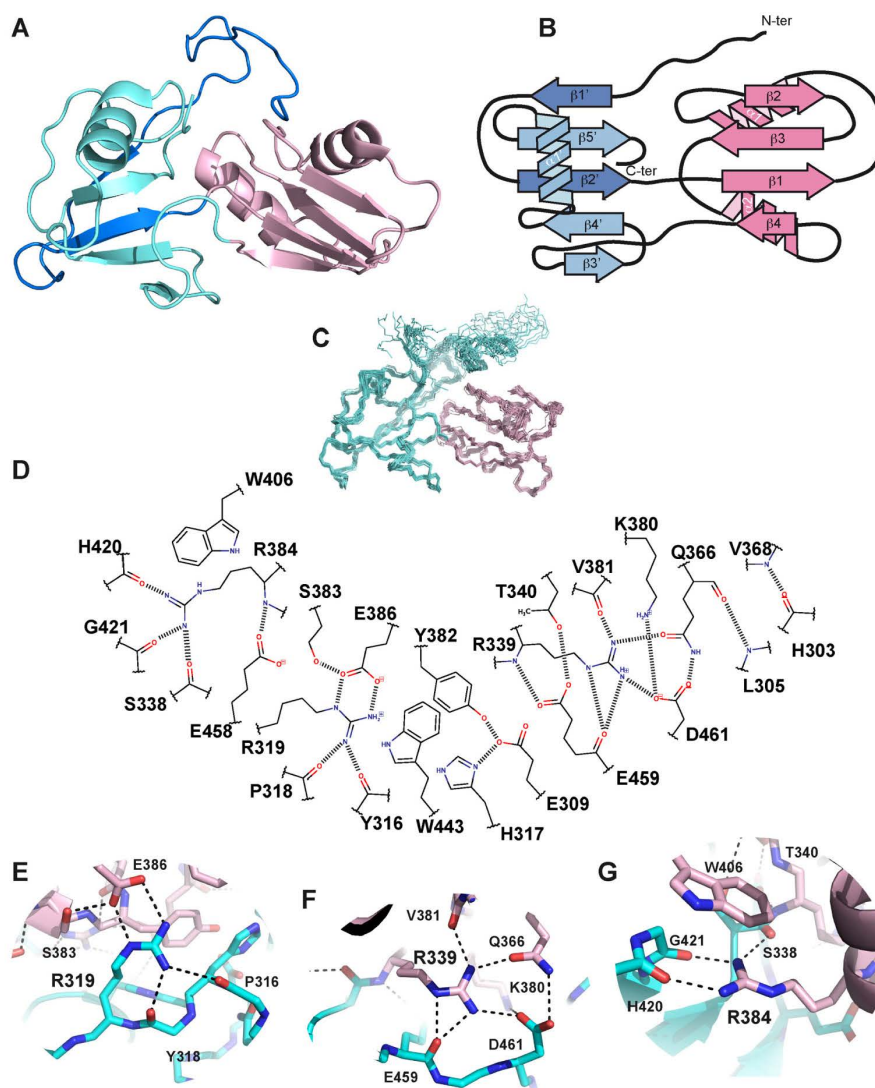


Figure 1. Structures of Nrd1 RNA binding domain. (A) Cartoon representation showing the RRM domain in pink and the SD in blue (different blue color is used for fragments non consecutive in sequence). (B) Schematic representation of the protein topology showing the elements of secondary structure. (C) Superposition of the NMR ensemble of the Nrd1₂₉₀₋₄₆₈. (D) Scheme of interactions formed at the RRM–SD interface. Hydrogen bonds are shown in dashed lines. (E) Structural detail around the Arg319 cluster. Residues are shown as sticks and colored in the same code as in A. Hydrogen bonds are shown in dashed lines. (F) Structural detail around the Arg339 cluster. (G) Structural detail around the Arg384 cluster.

that the heterogeneity in the longer construct might arise from slow exchange equilibrium of Phe298 coming in and out of the Leu305/Trp353 site. This hypothesis is compatible with the NOE data (Supplementary Figure S6D).

Altogether the X-ray and NMR data for Nrd1 RBD show the coexistence of a well-folded two-domain core with N-terminal and C-terminal tails that can interact loosely or in multiple ways with it.

Nrd1 RRM and SD domains interact through a conserved polar interface

The SD domain contains long loops with non-regular structure elements that interact with the upper loops of RRM

defining a large convex interface (Supplementary Figure S8A). This intramolecular interface is predominantly polar and maintained by an intricate network around three arginine residues (Arg319, Arg339 and Arg384; Figure 1D–G). On one side, Arg319 (partially exposed) forms a salt-bridge with Glu386 and hydrogen bonds with backbone carbonyls that freeze the orientation of the guanidium group (Figure 1E). At the most buried region of the interface, Arg339 coordinates an even more complex network of hydrogen bonds/salt bridges with Glu459, Asp461 and Gln366 (Figure 1F). A second basic residue (Lys380) takes part on this network and is probably important to keep this buried cluster neutral. A layer of aromatics (Trp443, Tyr382, Tyr306

and Trp353) separates the Arg339 and Arg319 clusters (Supplementary Figure S8B). Finally, Arg384 (partially exposed) closes the opposite end of the interface by a similar hydrogen bond network with backbone carbonyls of domain SD (Figure 1G). Despite its polar character, the SD–RRM interface is dry; only two water molecules (with low solvent accessibility and temperature factors) take part on it (Supplementary Figure S8C).

Nrd1 recognizes GUAA specifically

RRM domains typically recognize RNA using exposed residues in the β -sheet (47–49), a potential interface that is accessible in the structure of the Nrd1 RBD. Classical studies (10,12,13) and recent genome-wide CLIP maps (25,26) proposed Nrd1 selectivity for the sequences GUAA and GUAG. We have determined the binding affinities by ITC and fluorescence anisotropy showing that Nrd1_{290–468} recognizes GUAA with slightly higher affinity than GUAG and both of them with significantly larger affinity than the AUAA, AUAU and AUUA probes (Figure 2 and Table 1). The relatively high affinity of the Nrd1:GUAA interaction suggests that it might involve further residues than the canonical ones in RRM, because single RRM typically bind weaker than 2 μ M and without sequence specificity. To unravel this issue we attempted to obtain the structures of protein–RNA complexes by soaking and co-crystallization with short RNAs. Only the soaking method, and using tetragonal crystals from the Nrd1_{290–468} construct, was successful and we obtained three complex structures of Nrd1 with GUAA, CGUAAA and UUAGUAAUCC RNAs (Figure 3). The three sequences contain the GUAA motif, found in the majority of transcripts processed through the NNS pathway (12,25,26), and the last one derives from the *SNR13* terminator, one of the best characterized NNS targets (13). The core GUAA shows the same conformation in the three Nrd1:RNA complexes and there was enough electron density to build extra bases for the longer RNAs (Figure 3C). However these additional bases have higher temperature factors than the GUAA core and no evident contacts with the protein, thus we reasoned that they are loosely stabilized by intra RNA contacts. The Nrd1_{290–468}:GUAA crystal structure is consistent with NMR chemical shift perturbation data in solution, which map the RNA binding pocket to the depression between RRM and SD domains (Figure 3B). In conclusion, the structures suggest that Nrd1 RBD mainly recognises the core GUAA.

The Nrd1_{290–468}:GUAA structure, with the highest resolution among the three complexes (Supplementary Table S2), shows that both RRM and SD domains contribute to GUAA recognition (Figure 3). The RNA interacts with the RRM domain with its backbone running in the classical 5' to 3' orientation in all the RRM–RNA complexes reported to date. A series of hydrogen bonds (G1 O2'–His376 N δ 2, U2 O2'–A3 OP1 and A3 O2'–A4 O5') explain ribose specificity; in the case of U2, an unusual 2'-endo configuration for the sugar puckering favors this ribose-specific recognition (Figure 3D,F). The RRM domain interacts with the first three nucleotides: U2 (Figure 3F) and A3 (Figure 3G) form archetypal planar stacking interactions with Phe342 and Phe378 respectively, whereas G1 stacks to the edge of

Phe342 ring (Figure 3E). SD Residues Ile369 and Tyr418 make additional stacking interactions with A3 and U2, respectively. The fourth base (A4) forms a planar stacking with A3 (Figure 3H), which together with Phe378 define a three-layer aromatic array (Figure 3D&H). Residues from SD (His303, Ile462 and Val464) make further contacts with A4 (Figure 3H).

RNA base specificity is achieved by direct and water-mediated hydrogen bonds to the protein. Positions 3 and 4 are specified by direct hydrogen bonds between the N6 positions of the adenines and the backbone carbonyls of Val408 and Gly409 (Figure 3G and H). A4 N7 position (purine over pyrimidine selectivity) is specifically recognized by Arg413 and its mutation lowers affinity \sim 10-fold (Table 1), reinforcing the structural role of this conserved residue (Supplementary Figure S3) in purine recognition. A3 N7 position is recognized by a water molecule that contacts simultaneously to U2 O2. The remaining acceptors/donors of A3 and A4 are satisfied by interactions with four water molecules that are trapped between the RNA and the protein (Figure 3G and H). These waters are well structured, according to their temperature factors, but are not completely buried by the interaction; a narrow channel on the back of this water pocket communicates it with the bulk solvent (Supplementary Figure S9). U2 position is specified by the U2 N3–Trp406 O and by the U2 O2–water hydrogen bonds (Figure 3F). The conserved Tyr418 interacts with U2 and its mutation decreases RNA affinity (Table 1). Finally, two hydrogen bonds (G1 N2–U2 OP1 and G1 N7–Gly345 NH) specify guanine at position 1 (Figure 3E). There is a third base-specific contact with a Tris molecule (from the buffer) that bridges G1 with the Arg403/Arg405 pair. The Nrd1_{290–468} crystal structure in complex with CGUAAA, which was solved in phosphate buffer (no Tris present), showed that only Arg403 makes base-specific contacts (Arg403 N ϵ –G1 O6) (Supplementary Figure S10). This result is consistent with mutational data on the equivalent residue in Seb1 which lowers RNA affinity dramatically (46). To enquire about the role of Arg405 in RNA recognition we measured energetics of binding of the R405A mutant finding slightly higher affinity than for *wt* protein (Table 1). Therefore, we confirm Arg403 as the one involved in G1 O6 recognition. To further confirm the guanine requirement at position 1, we tested substitution by other purines. Substantial affinity losses were observed upon inosine (\sim 15-fold: loss of N3 group) and adenine ($>$ 40-fold: full remodeling of the Watson–Crick face) substitutions (Table 1 and Figure 2C).

Structural integrity of the RRM–SD tandem is key for RNA recognition

RNA-binding causes little structural changes on Nrd1 (Supplementary Figure S11), suggesting that the RNA binding site is structurally preconfigured. The X-ray structures of the different protein–RNA complexes reveal the key elements for the RNA recognition. We made a thorough mutagenesis analysis to gain insight into different contributions to RNA binding energetics.

R374A, R413G and Y418A show a rather homogeneous decrease in affinity (15-fold K_D increase) (Table 1),

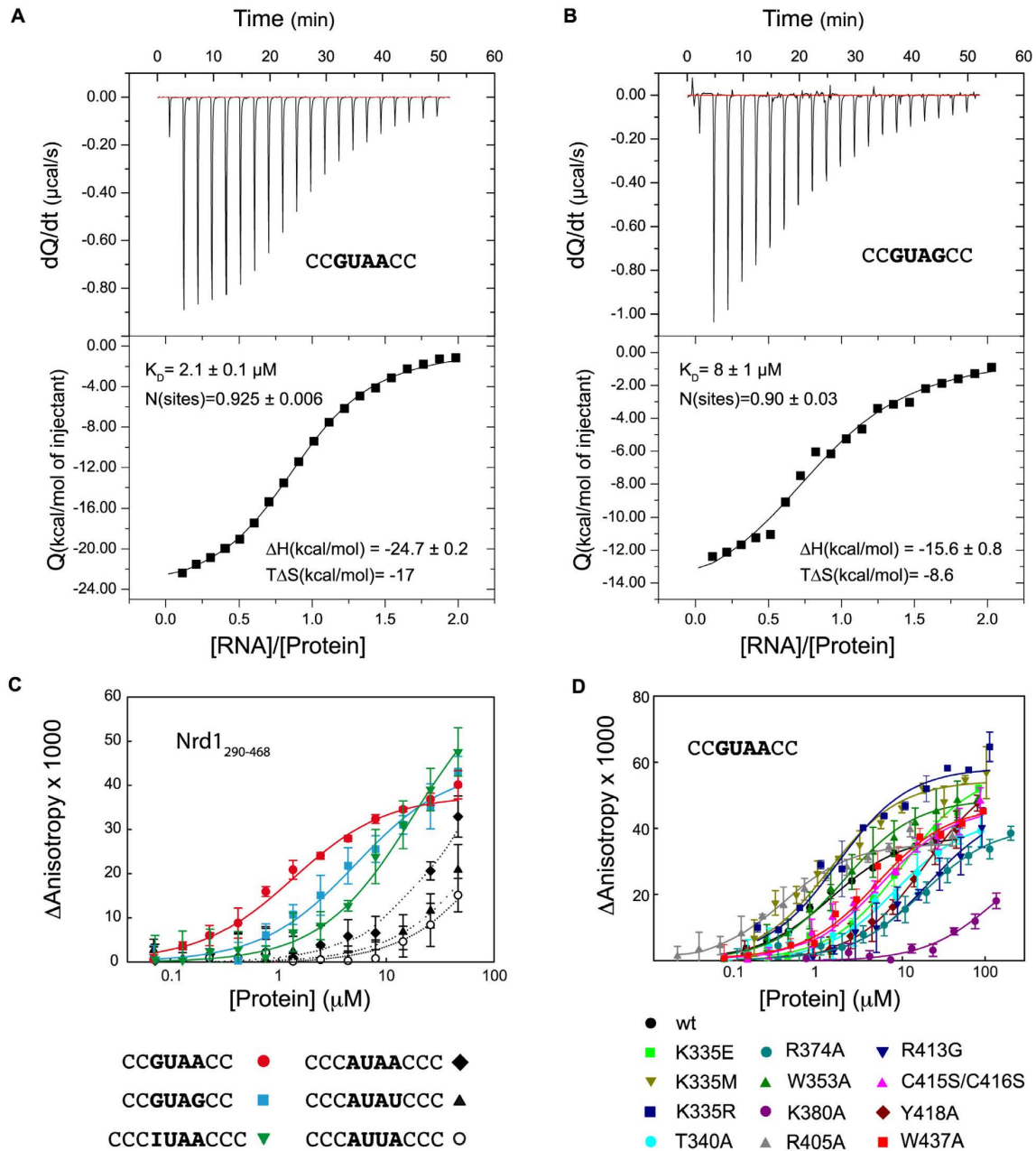


Figure 2. Biophysical analysis of RNA recognition by Nrd1 RBD. (A and B) Isothermal titration calorimetry thermograms (upper panels) and binding isotherms (lower panels) obtained for titrations of CCGUAACC and CCGUAGCC RNAs over Nrd1₂₉₀₋₄₆₈. Enthalpic and entropic contributions to binding, K_D ($1/K_B$) and stoichiometry (N) are indicated on each curve. (C) Fluorescence anisotropy Nrd1₂₉₀₋₄₆₈ titration curves over different RNAs (I = ionosine). (D) Fluorescence anisotropy CCGUAACC titration curves over various Nrd1₂₉₀₋₄₆₈ mutants. Fitted values are summarized in Table 1.

presumably due to loss of protein–RNA contacts (hydrogen bonds: R413G, hydrophobic: Y418A and columbic: R374A). We also performed mutations in the second coordination sphere (stabilizes residues in direct contact with RNA): T340A, W437A and C415S/C416S, exhibiting more modest effects (Table 1). Next we altered the RRM–SD interface with different results: W353A mutant, that should

affect the interaction of the N-terminus with the RRM, has a very little impact on GUAA recognition. In stark contrast, K380A, at the heart of the RRM–SD interface, lowers RNA affinity dramatically, likely due to a disruption of the interdomain arrangement. Finally, we mutated the conserved residue Lys335 (Supplementary Figure S3), placed at the centre of the SD β -sheet and far away from

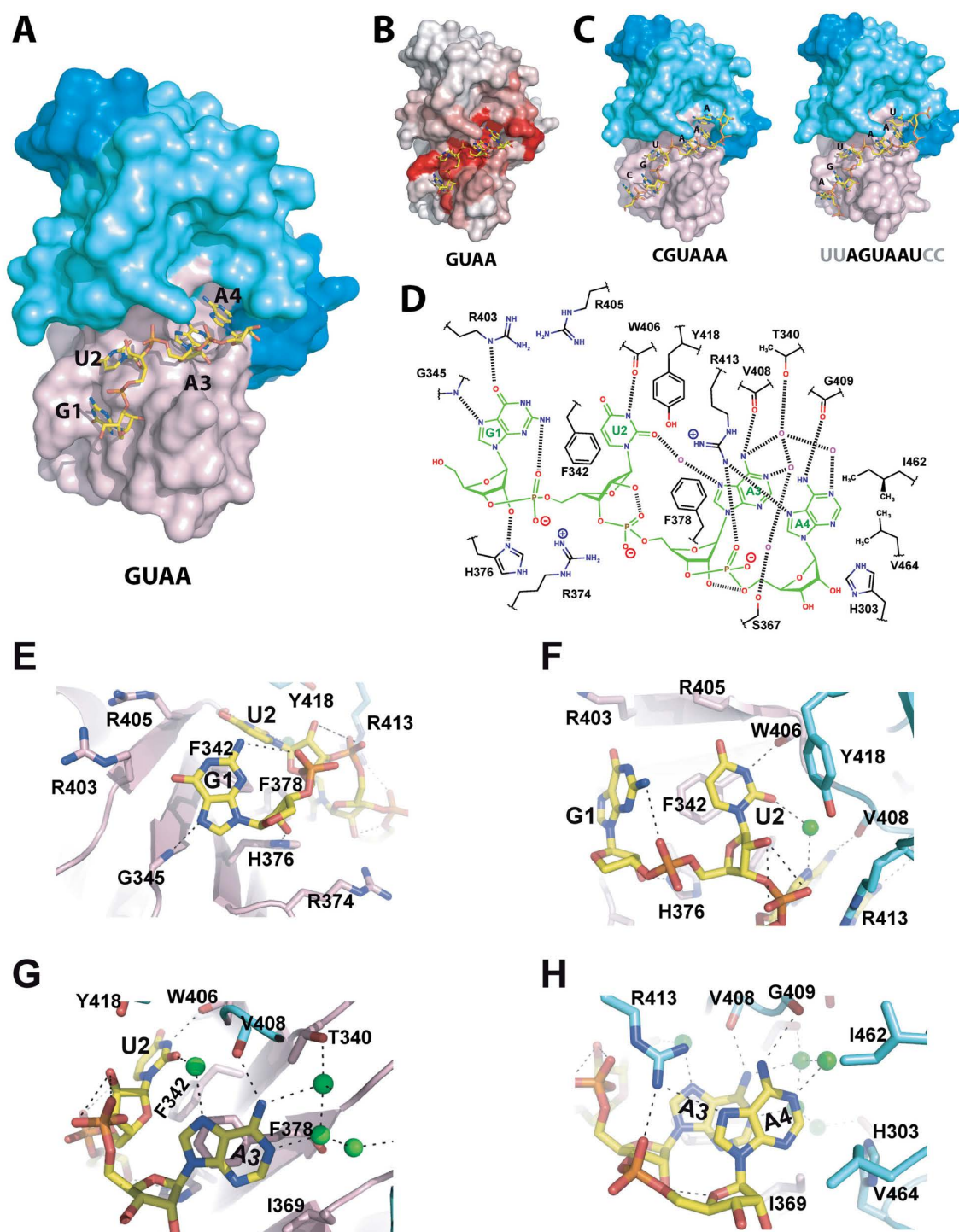


Figure 3. X-ray structures of the Nrd1₂₉₀₋₄₆₈:RNA complexes. (A) Nrd1₂₉₀₋₄₆₈:GUAA complex represented as surface and colored by domains as in Figure 1. The RNA is shown as sticks (C: yellow, O: red, N: blue and P: orange). (B) Nrd1₂₉₀₋₄₆₈:GUAA complex in the same orientation as A and with the surface colored by chemical shift mapping obtained for the same interaction by NMR (shades of red). (C) Nrd1₂₉₀₋₄₆₈:CGUAAA (left) and Nrd1₂₉₀₋₄₆₈:UUAGUAAUCC (right) complexes. (D) Schematic representation of the Nrd1:GUAA interface. Hydrogen bonds are represented in dashed lines. (E) Structural detail around G1. (F) Structural detail around U2. (G) Structural detail around A3. (H) Structural detail around A4. Water molecules are shown as green spheres. Side-chains of residues interacting through their backbone carbonyls or amide groups have been omitted for clarity.

Table 1. Thermodynamic parameters of *wt* and mutants Nrd1_{290–468}/RNA interactions determined by isothermal titration calorimetry (ITC) and fluorescence anisotropy (FA)

Exp.	Protein	Mutant	RNA ^a	ΔG (kcal/mol) ^b	K_D (μ M) ^c
ITC	Nrd1	wt	CCGUAACC	7.74 ± 0.05	2.1 ± 0.1
ITC	Nrd1	wt	CCGUAGCC	6.95 ± 0.08	8 ± 1
FA	txA-Nrd1	wt	CCGUAACC	8.0 ± 0.1	1.4 ± 0.2
FA	txA-Nrd1	wt	CCGUAGCC	7.2 ± 0.2	5 ± 1
FA	txA-Nrd1	wt	CCCAUAACCC	n.d.	$\sim 50^d$
FA	txA-Nrd1	wt	CCCIUAACCC	6.5 ± 0.2	18 ± 5
FA	txA-Nrd1	wt	CCCAUUAACCC	n.d.	$\sim 160^d$
FA	txA-Nrd1	wt	CCCAUAUCC	n.d.	$\sim 110^d$
FA	txA-Nrd1	K335E	CCGUAACC	6.8 ± 0.1	11 ± 1
FA	txA-Nrd1	K335M	CCGUAGCC	8.1 ± 0.1	1.2 ± 0.2
FA	txA-Nrd1	K335R	CCGUAACC	7.9 ± 0.2	1.7 ± 0.4
FA	txA-Nrd1	T340A	CCGUAGCC	6.9 ± 0.2	9 ± 2
FA	txA-Nrd1	W353A	CCGUAACC	7.8 ± 0.1	2.0 ± 0.3
FA	txA-Nrd1	R374A	CCGUAGCC	6.5 ± 0.2	18 ± 5
FA	txA-Nrd1	K380A	CCGUAGCC	n.d.	$\sim 135^d$
FA	txA-Nrd1	R405A	CCGUAGCC	8.8 ± 0.1	0.37 ± 0.05
FA	txA-Nrd1	R413G	CCGUAGCC	6.4 ± 0.2	21 ± 6
FA	txA-Nrd1	C415S/C416S	CCGUAGCC	7.2 ± 0.2	5 ± 1
FA	txA-Nrd1	Y418A	CCGUAGCC	6.4 ± 0.1	21 ± 3
FA	txA-Nrd1	W437A	CCGUAGCC	7.3 ± 0.2	5 ± 1

^aSequences 5' to 3'. All RNAs used for fluorescence anisotropy were fluorescein-labeled at 5'. I = inosine.

^bDissociation ΔG obtained in FA from non-linear least squares fitting of the data. ITC values and uncertainties were calculated from $\Delta G = RT \ln(K_{B,obs})$ and error propagation.

^c $K_D = (K_{B,obs})^{-1}$ in ITC. Apparent K_D values and uncertainties in FA were calculated from $K_D = e^{(-\Delta G/RT)}$ and error propagation.

^dEstimations assuming similar fluorescence anisotropy changes as those obtained in the curves in which saturation was reached.

n.d.: not determined due to insufficient saturation.

the RNA-binding site (Supplementary Figure S11). This is not involved in RNA recognition, however the K335E exhibits a small GUAA affinity loss (Table 1). Lys335 forms a conserved salt bridge with Asp326, thus we argue that the K335E mutant might introduce a destabilizing repulsion effect. To prove this, we measured RNA binding affinity of more structurally compatible K335R and K335M mutants, which rendered values indistinguishable from *wt*.

Altogether, we conclude that the SD plays a scaffolding role allowing the structural preconfiguration of the GUAA binding site in Nrd1.

The split domain of Nrd1 is essential for viability

Genomic removal of Nrd1 compromises cell viability (10). We aimed to find out if the Nrd1 SD (Figure 4A) affects cell survival by introducing LEU plasmids containing wild-type *NRD1* (*wt*), complete SD deletion ($\Delta 301-336/\Delta 412-463$) or partial deletions ($\Delta SD_A = \Delta 301-336$ and $\Delta SD_B = \Delta 412-463$) of this domain in a yeast strain harbouring a centromeric URA plasmid expressing a genomic copy of *NRD1* (10). Loss of the URA wild-type plasmid upon 5-FOA treatment and plasmid shuffling results in cell lethality in all the SD deletion: only the cells bearing *wt NRD1* remain viable (Figure 4B left panel). We further tested synthetic lethality of these SD deletion mutants using another method: controlling genomic *wt NRD1* expression under the GAL promoter. As shown, cells harboring SD deletion mutations are unable to grow in glucose containing media (Figure 4B, right panel). Therefore, the SD is a key element for Nrd1 activity due to either its role in specific recognition of RNA terminators (as described in our Nrd1:RNA structures) or to other unknown functions. We extended our mu-

tagenesis study to delve deeper into these important questions.

Nrd1 has a conserved sequence at the C-terminus (residues 551–575) (Supplementary Figure S12), which includes a polyglutamine tail. We introduced stop codon mutations at several positions after the RBD (L558stop, Q567stop and K490stop) that do not cause cell growth defects (Figure 4C), suggesting that the conserved Nrd1 C-terminus is not critical for its function. In contrast, disrupting Nrd1 from the Val468 residue (V468stop) provokes significant growth defects at all tested temperatures (Figure 4C, upper panel). The *nrd1* $\Delta 301-306$ mutant, at the N-terminus of the RBD, shows less marked growth defects (Figure 4C, bottom panel).

Next we explored how changes in RNA affinity affect Nrd1 function *in vivo*. In general, mutations causing minor affinity losses (W353A, C415S/C416S, W437A) do not display noticeable growth defects. Although in the R374A, R413G and Y418A mutants, affecting direct Nrd1–RNA contacts and displaying similar K_D increases (Table 1), only Y418A cells present slow growth at 37°C (Figure 4C, bottom panel). Conversely, K335E, with less impact in RNA recognition, has a strong growth defect, which is comparable with the previously reported defects on T340A and K380A (24) (Figure 4C). More importantly, Y418A/K335E double mutation (with both residues belonging to the SD) causes an extraordinary slow growth at 28°C and lethality at 37°C (Figure 4C). Therefore, this double mutation is almost as aggressive as the SD deletion mutants (Figure 4B). The growth defect caused by the K335E is not totally rescued in the case of K335M and K335R (Figure 4C). These two mutants show RNA binding affinities indistinguishable from that of *wt* (Table 1), therefore we

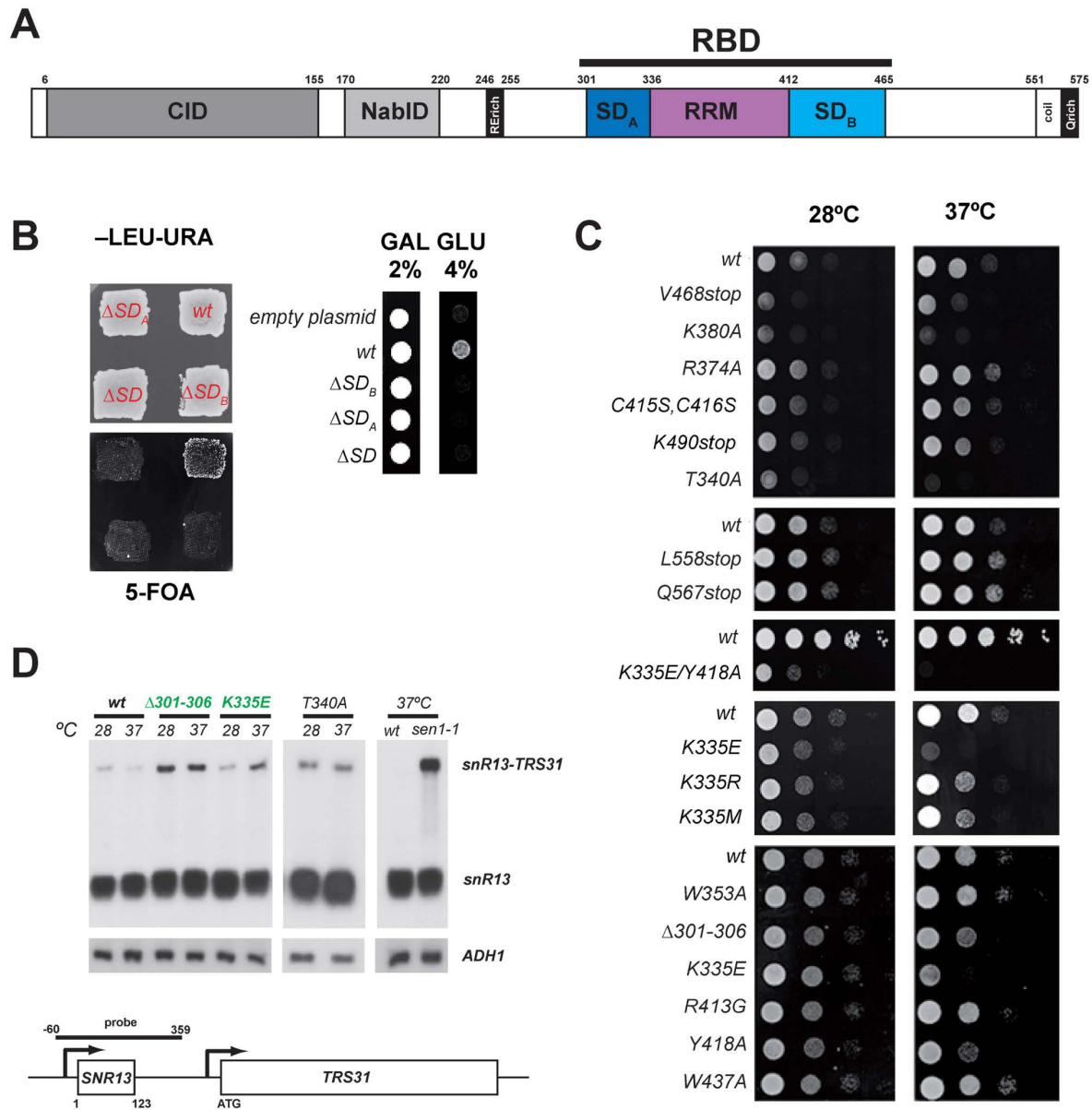


Figure 4. Nrd1 RBD functional analysis. (A) Schematic representation of Nrd1 domain organization. (B) Partial or total deletion of the Nrd1-SD causes cell lethality. Left panel, the indicated strains were grown in -LEU-URA selective media and then replicated onto 5-FOA containing media and grown for 2–3 days at 28°C. Right panel, strains with the indicated genotypes were grown in galactose (GAL) or glucose (GLU) containing media and grown for 2–3 days. (C) Analysis of *nrd1* mutants growth phenotypes. Serial dilutions (1:10) of *wt* and *nrd1* mutant strains were spotted on selective SC media and grown for 2–3 days at the indicated temperatures. (D) Northern blot analyses of the *SNR13* and *ADH1* genes. Total RNA was purified from the indicated strains grown at 28°C and 37°C. In addition, *wt* and *sen1-1* isogenic strains were used as positive control for transcription termination defects (50). Full-length snR13 and readthrough transcripts are indicated by arrows. A schematic representation of *SNR13* gene and 3' end flanking region is also shown, where the size and localization of the probe is represented by a black bar above *SNR13* coding region.

propose that they alter some functions of Nrd1 RBD not related to their RNA binding; probably unknown protein-protein interactions with other components of the NNS pathway.

Finally we decided to test whether the mutants exhibiting reduced cell growth phenotypes also showed transcription termination defects at both 28°C and 37°C. For that purpose, we analysed the well-known *SNR13* transcript by northern blot assay as described in (50). We used as a positive control of snoRNA transcription termination defects the *sen1-1* mutant grown at 37°C (50). *SNR13* displayed strong transcription termination defects at 28°C and 37°C in *nrd1* Δ 301–306 mutant, and to a lesser extent, but significantly, in *nrd1*-K335E cells at 37°C, consistent with the extraordinary reduced growth of that mutant at this temperature. In the case of *nrd1*-T340A cells, whose mutation lies into the RRM domain, *SNR13* transcription termination defects are also observed at both temperatures (Figure 4D). Therefore, these data suggest that both RRM and SD have a role in snoRNAs transcription termination. The fact that K335E and T340A mutants do not show a very marked RNA affinity lost (Table 1) lead us to think the biological role of Nrd1 RBD in snoRNAs transcription termination might not be solely dependent on RNA binding.

DISCUSSION

Widely found in eukaryotes, the RRM has a typical architecture (β_1 – α_1 – β_2 – β_3 – α_2 – β_4) in which the RNA recognition residues (RNP1 and RNP2 motifs) are on the exposed face of β_2 and β_3 (47–49,51,52). From this basic fold, there are a number of RRM with extended features, typically built from additional regions on the C- and/or N-termini: extra β -strands (53), α -helices (54–56) and β -sheets (27). The structure of Nrd1 RBD is a new class of RRM; it contains additional N- and C-terminal extensions which, rather than adding extra secondary structural elements, form together a domain by itself (SD). The SD is fused to the RRM through an intricate polar interface, defining a rigid body structure. The structures of Nrd1 and Seb1 (46) RBDs have similar folds, differing in that Nrd1 RBD includes long N-terminal extensions that make further contacts with the RRM domain (Supplementary Figure S13 A&B). The RRM–SD interface is mostly conserved between Nrd1 and Seb1 RBDs (Supplementary Figure S13C–E). Only the Arg319 cluster is not present in Seb1 and the rest of RRM–SD interfacial residues: Arg339, Lys380, Arg384, Gln366, Trp406, Pro460 and Asp461, are totally conserved in fungi (Supplementary Figure S3), suggesting a conservation of the Nrd1/Seb1 RBD structure at least in yeast and moulds.

Seb1 RBD binds AUUAGUAAAA with 1.8 μ M affinity (46), a remarkably close value to the 1.5–2.1 μ M affinity that we measured for Nrd1 RBD binding to CCGUAACC. Both Seb1 and Nrd1 RNA targets contain the GUAA motif. These similarities suggest that Seb1 might have a similar RNA binding mode than Nrd1, hence we constructed a model of the Seb1:RNA complex by placing the RNA on the Seb1 structure in the same conformation as in the Nrd1_{290–468}:GUAA complex (Supplementary Figure S13F). This simple model is likely realistic because, as it happens

for Nrd1 (Supplementary Figure S11), we expect little conformational changes in Seb1 upon binding. In support of this model, the mutagenesis data (46) show that the biggest changes in RNA binding affinity affect interfacial residues, while other mutants, far from the hypothetical binding site, cause negligible effects. In conclusion, we predict that Seb1 will bind RNAs with the GUAA/G motif by using a similar mode as Nrd1 RBD, which is probably conserved across the family with minor variations.

Both GUAA and GUAG sequences have been identified at NNS terminator regions (10,12,13). Nrd1 binds GUAG terminators with four times lower affinity than for GUAA ones. In both cases binding is enthalpically driven (Figure 2) in agreement with the formation of multiple polar interactions: hydrogen bonds and salt bridges. However, the magnitude of ΔH is lower for the GUAG complex (Figure 2) suggesting that there are less polar contacts in this case. In the Nrd1:GUAA complex, A4 acts as a sort of molecular gate trapping four water molecules between the RNA and the SD (Figure 3G). We hypothesize that a suboptimal binding of G4 would facilitate the release of these ligand-binding waters providing an explanation for the loss in ΔH contribution in GUAG complex formation with respect to that of GUAA one.

The uniqueness of Nrd1-like RBD structure opens the question of whether this domain might play further roles beyond RNA binding. We demonstrated that the SD accompanying the canonical RRM is important for cell viability (Figure 4B) and its role in RNA binding is demonstrated by structural and mutagenesis data. SD mutants *nrd1*-K335E and *nrd1* Δ (301–306) exhibit both cell growth and *SNR13* termination defects (Figure 4C and D). In the case of Lys335, a conserved residue in the opposite face of the RNA binding pocket, its mutation might slightly perturb the SD stability/architecture but not sufficiently to cause a dramatic change in RNA binding affinity (Table 1). Instead the *nrd1*-K335E phenotype is likely due to the disruption of other interactions within the NNS pathway. Currently, we have sufficient structural/biochemical information to designate potential targets for Nrd1 RBD protein–protein interactions. The two known Nrd1 anchoring points to the RNA pol II are the nascent RNA and the pS5 Rpb1 CTD repeats, both facing toward the same side of the enzyme (Supplementary Figure S14). Nrd1 interacts with Nab3 through its heterodimerization domain (14) and both co-purify with the RNA helicase Sen1 (57). However, Sen1 recruitment to the NNS complex seems to be mediated by Nab3, rather than by Nrd1 (58). On the other hand, the RNA Pol II stalk, formed by the Rpb4/Rbp7 heterodimer, participates in the recruitment of RNA processing factors and CTD modifying enzymes (59–61). Interestingly, nascent RNA exits making contacts with Rpb7 (62) and Nrd1 interacts physically with Rpb7 through its C-terminus (including half of the SD) (63). The temperature sensitive phenotype of *nrd1*-V468stop might be due to a partial or total disruption of this interaction. If so, the Nrd1-Rpb7 interaction would locate the Nrd1 RBD on the direct exit route of the newly synthesized RNA, allowing an early recognition of GUAA/G sequences, which perhaps triggers transcription termination.

ACCESSION NUMBERS

Atomic coordinates of the protein structures and their complexes have been deposited in the Protein Data Bank (PDB) with the following accession codes: 5O1W (crystal structure of Nrd1_{301–489}); 5O1X (crystal structure of Nrd1_{290–468}); 5O1Y (crystal structure of Nrd1_{290–468} in complex with GUAA RNA); 5O1Z (crystal structure of Nrd1_{290–468} in complex with CGUAAA RNA); 5O20 (crystal structure of Nrd1_{290–468} in complex with UUAGUAAUCC RNA (corresponds to SNR13 148–157)) and 5O1T (Nrd1_{290–468} NMR structure). The chemical shifts for Nrd1_{290–468} construct has been deposited on the Biological Magnetic Resonance Bank (BMRB) with the accession code 34140.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr S. Buratowski for gratefully providing yeast strains and plasmids. Data collection was performed at ESRF, ID23-1 beamline (Grenoble, France) and at ALBA Synchrotron, XALOC-BL13 beamline, with the collaboration of ESRF and ALBA staff.

FUNDING

Spanish Ministry of Economy and Competitiveness (MINECO) [BFU2014-53762-P to B.G., BFU2015-71978-REDT and BFU2013-48374-P to O.C., CTQ2011-26665 and CTQ2014-52633 to J.M.P.C.]; E.F.-E. was supported by BFU2014-53762-P grant. Funding for open access charge: MINECO.

Conflict of interest statement. None declared.

REFERENCES

- Tudek, A., Candelli, T. and Libri, D. (2015) Non-coding transcription by RNA polymerase II in yeast: Hasard or necessity? *Biochimie*, **117**, 28–36.
- Porrua, O., Boudvillain, M. and Libri, D. (2016) Transcription Termination: Variations on Common Themes. *Trends Genet.*, **32**, 508–522.
- Loya, T.J. and Reines, D. (2016) Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research*, **5**, F1000.
- Porrua, O. and Libri, D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.*, **16**, 190–202.
- Fasken, M.B., Larabee, R.N. and Corbett, A.H. (2015) Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rps6 independent of Nrd1. *PLoS Genet.*, **11**, e1005044.
- Arndt, K.M. and Reines, D. (2015) Termination of transcription of short noncoding RNAs by RNA polymerase II. *Annu. Rev. Biochem.*, **84**, 381–404.
- Steinmetz, E.J., Conrad, N.K., Brow, D.A. and Corden, J.L. (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, **413**, 327–331.
- Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. and Libri, D. (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the Nrd1-Nab3 pathway in genome surveillance. *Mol. Cell*, **23**, 853–864.
- Wilson, S.M., Datar, K.V., Paddy, M.R., Swedlow, J.R. and Swanson, M.S. (1994) Characterization of nuclear polyadenylated RNA-binding proteins in *Saccharomyces cerevisiae*. *J. Cell Biol.*, **127**, 1173–1184.
- Steinmetz, E.J. and Brow, D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.*, **16**, 6993–7003.
- Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R. and Libri, D. (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.*, **31**, 3935–3948.
- Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D. and Corden, J.L. (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol. Cell. Biol.*, **24**, 6241–6252.
- Carroll, K.L., Ghirlando, R., Ames, J.M. and Corden, J.L. (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA*, **13**, 361–373.
- Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S. and Corden, J.L. (2000) A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics*, **154**, 557–571.
- Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. and Meinhart, A. (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, **15**, 795–804.
- Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S. and Stefl, R. (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.*, **26**, 1891–1896.
- Grzechnik, P. and Kufel, J. (2008) Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol. Cell*, **32**, 247–258.
- Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R. *et al.* (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell*, **55**, 467–481.
- Loya, T.J., O'Rourke, T.W. and Reines, D. (2012) Yeast Nab3 Contains a Self-Assembly Domain found in Human hnRNP-C that is Necessary for Transcription Termination. *J. Biol. Chem.*, **288**, 2111–2117.
- Meinhart, A. and Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, **430**, 223–226.
- Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A. and Varani, G. (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.*, **17**, 1195–1201.
- Lunde, B.M., Horner, M. and Meinhart, A. (2011) Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.*, **39**, 337–346.
- Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S. and Stefl, R. (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.*, **286**, 3645–3657.
- Bacikova, V., Pasulka, J., Kubicek, K. and Stefl, R. (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.*, **42**, 8024–8038.
- Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J. and Corden, J.L. (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA*, **17**, 2011–2025.
- Schulz, D., Schwalb, B., Kiesel, A., Baerjen, C., Torkler, P., Gagneur, J., Soeding, J. and Cramer, P. (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*, **155**, 1075–1087.
- Martinez-Lumbreras, S., Taverniti, V., Zorrilla, S., Séraphin, B. and Pérez-Cañadillas, J.M. (2016) Gbp2 interacts with THO/TREX through a novel type of RRM domain. *Nucleic Acids Res.*, **44**, 437–448.

28. Bergfors, T. (2003) Seeds to crystals. *J. Struct. Biol.*, **142**, 66–76.
29. Kabsch, W. (2010) Xds. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 125–132.
30. Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D, Biol. Crystallogr.*, **67**, 282–292.
31. Sheldrick, G.M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 479–485.
32. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D, Biol. Crystallogr.*, **62**, 1002–1011.
33. Vagin, A. and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 22–25.
34. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D, Biol. Crystallogr.*, **66**, 486–501.
35. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D, Biol. Crystallogr.*, **53**, 240–255.
36. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
37. DeLano, W.L. (2002). *DeLano Scientific*, San Carlos, CA.
38. Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93–158.
39. Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ¹H and ¹³C correlations in H₂O samples of proteins. *J. Magn. Reson. Ser. B*, **101**, 333–337.
40. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
41. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
42. Breeze, A.L. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Mag. Res. Sp.*, **36**, 323–372.
43. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
44. Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, **273**, 283–298.
45. Bruke, D., Dawson, D. and Stearns, T. (2000) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, NY.
46. Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.H., Kecman, T. et al. (2017) The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat. Commun.*, **8**, 14861.
47. Daubner, G.M., Clery, A. and Allain, F.H. (2013) RRM-RNA recognition: NMR or crystallography... and new findings. *Curr. Opin. Struct. Biol.*, **23**, 100–108.
48. Clery, A., Blatter, M. and Allain, F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
49. Maris, C., Dominguez, C. and Allain, F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
50. Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C.L. and Buratowski, S. (2006) Distinct pathways for snoRNA and mRNA termination. *Mol. Cell*, **24**, 723–734.
51. Pérez-Cañadillas, J.M. and Varani, G. (2001) Recent advances in RNA-protein recognition. *Curr. Opin. Struct. Biol.*, **11**, 53–58.
52. Messias, A.C. and Sattler, M. (2004) Structural basis of single-stranded RNA recognition. *Acc. Chem. Res.*, **37**, 279–287.
53. Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. et al. (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054–2057.
54. Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S. and Conte, M.R. (2003) Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure*, **11**, 833–843.
55. Pérez-Cañadillas, J.M. and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.*, **22**, 2821–2830.
56. Santiveri, C.M., Mirassou, Y., Rico-Lastres, P., Martínez-Lumbreras, S. and Pérez-Cañadillas, J.M. (2011) Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. *PLoS One*, **6**, e24481.
57. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239–248.
58. Nedeá, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F. and Nagy, P.L. (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol. Cell*, **29**, 577–587.
59. Allepuz-Fuster, P., Martínez-Fernández, V., Garrido-Godino, A.I., Alonso-Aguado, S., Hanes, S.D., Navarro, F. and Calvo, O. (2014) Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation. *Nucleic Acids Res.*, **42**, 13674–13688.
60. Garavis, M., González-Polo, N., Allepuz-Fuster, P., Louro, J.A., Fernández-Tornero, C. and Calvo, O. (2017) Sub1 contacts the RNA polymerase II stalk to modulate mRNA synthesis. *Nucleic Acids Res.*, **45**, 2458–2471.
61. Runner, V.M., Podolny, V. and Buratowski, S. (2008) The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell Biol.*, **28**, 1883–1891.
62. Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T. and Chang, W.H. (2009) Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 127–132.
63. Mitsuzawa, H., Kanda, E. and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.*, **31**, 4696–4701.

SUPPLEMENTARY FILES

The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition

Elsa Franco-Echevarría, Noelia González-Polo, Silvia Zorrilla, Santiago Martínez-Lumbreras, Clara M Santiveri, Ramón Campos-Olivas, Mar Sánchez Olga Calvo, Beatriz González*, José Manuel Pérez-Cañadillas*.

TABLE OF CONTENTS

Supplementary Table S1	Plasmids and protein constructs used in this work
Supplementary Table S2	Crystallographic data statistics and refinement
Supplementary Table S3	Summary of NMR restrains and structural calculation statistics for Nrd1 ₂₉₀₋₄₆₈ solution structure.
Supplementary Table S4	NOE-derived distance restraints between RRM and SD domains
Supplementary Figure S1	Selectively unlabelled samples in ¹³ C/ ¹⁵ N background greatly facilitates NOEs assignments
Supplementary Figure S2	Selected NMR data and derived distance restraints
Supplementary Figure S3	Sequence alignment of RNA binding domains of 39 Nrd1 orthologues in the OMA group 108412
Supplementary Figure S4	Comparison of ¹ H- ¹⁵ N HSQC NMR spectra of Nrd1 ₃₀₁₋₄₈₉ and Nrd1 ₂₉₀₋₄₆₈ .
Supplementary Figure S5	Chemical Shift Index by residue for Nrd1 ₂₉₀₋₄₆₈
Supplementary Figure S6	Nrd1 RBD structures show conformational differences at their termini.
Supplementary Figure S7	Comparison between X-ray and NMR structures of Nrd1 RBD.
Supplementary Figure S8	Nrd1 RRM and SD subdomains interface details
Supplementary Figure S9	The structure of the Nrd1-RNA complex contains an internal water pocket
Supplementary Figure S10	Structure of Nrd1 ₂₉₀₋₄₆₈ in complex with CGUAAA obtained in phosphate buffer
Supplementary Figure S11	The Nrd1 RBD experiences minor changes upon GUAA recognition
Supplementary Figure S12	Conservation of C-terminal elements of Nrd1 in different yeasts
Supplementary Figure S13	Structural comparison between Nrd1 and Seb1 (PDB: 5MD4) RNA binding domains.
Supplementary Figure S14	Structural model of the NNS complex and elongating RNA polymerase.

Supplementary Table S1: Plasmids and protein constructs used in this work. Keywords: NMR=Nuclear Magnetic Resonance, X-ray= X-ray crystallography, F.A.= Fluorescence anisotropy, ITC= Isothermal titration calorimetry, Fun. stu.= Functional studies in *S. cerevisiae* (drop tests and northern block). pET28 txAHTEV is a pET28-derived plasmid in which the 6xHis-Trombin site has been replaced by trxA-6xHis-TEVsite

Plasmid/vector	Expression host/strain	Use	Source
pET28-txAHTEV-Nrd1 301-489	<i>E. coli</i> BL21(DE3)	X-ray; NMR:F.A.	This work
pET28-txAHTEV-Nrd1 265-489	<i>E. coli</i> BL21(DE3)	NMR	This work
pET28-txAHTEV-Nrd1 290-468	<i>E. coli</i> BL21(DE3)	X-ray; NMR;F.A., ITC.	This work
pRS316-NRD1	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3 [pRS316-NRD1]</i>	Fun. stu.	(1-3)
pRS415-NRD1	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3 [pRS415-NRD1]</i>	Fun. stu.	This work
pRS415-nrd1-Δ301-306	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-Δ301-336	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-Δ412-463	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-Δ301-336,Δ412-463	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-V467stop	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-K490stop	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-L558stop	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-Q567stop	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-K335E	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-K335R	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-K335M	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work

Supplementary Table S1 (cont.)

Plasmid/vector	Expression host/strain	Use	Source
pRS415-nrd1-T340A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-W353A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-R374A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-K380A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-R413G	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-Y418A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-W437A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1- K335E,Y418A	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pRS415-nrd1-C415S,C416S	<i>Mat a ura3-52, leu2-3,112, trp1-1, his3-11,15, ade2-1, met2Δ1, lys2Δ2, can1-100, nrd1ΔHIS3</i>	Fun. stu.	This work
pET28-txAHTEV-Nrd1 290-468 K335E	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K335M	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K335R	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 T430A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 W353A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R374A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 K380A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R405A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 R413G	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 Y418A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 W437A	<i>E. coli</i> BL21(DE3)	F.A.	This work
pET28-txAHTEV-Nrd1 290-468 C415S/C416S	<i>E. coli</i> BL21(DE3)	F.A.	This work

Supplementary Table S2: Crystallographic data statistics and refinement.

Dataset	Nrd1 301-488 SeMet	Nrd1 301-488	Nrd1 290-468	Nrd1 290-468	Nrd1 290-468	Nrd1 290-468
Data collection						
RNA	-	-	-	GUAA	CGUAAA	UUAGUAAUCC
RNA built	-	-	-	GUAA	CGUAAA	AGUAAU
Space group	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P6 ₅	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2
Unit cell, <i>a,b,c</i> (Å)	61.05, 61.05, 157.60	60.99, 60.99, 157.96	58.39, 58.39, 92.45	64.94, 64.94, 158.39	64.53, 64.53, 157.03	66.16, 66.16, 156.91
γ (Å)			120.0			
Synchrotron	XALOC (ALBA)	ID23-2 (ESRF)	XALOC (ALBA)	XALOC (ALBA)	XALOC (ALBA) PILATUS	XALOC (ALBA) PILATUS
Beamline/Detector	PILATUS 6M	PILATUS 2M	PILATUS 6M	PILATUS 6M	6M	6M
Temperature (°C)	-173	-173	-173	-173	-173	-173
Wavelength (Å)	0.979330	0.976250	0.97930	0.979260	0.979290	0.979260
Resolution (Å)	48.26 - 2.20 (2.27-2.20)	48.14 - 2.30 (2.38-2.30)	92.45 - 1.60 (1.63-1.60)	158.39 - 2.45 (2.55 -2.45)	45.63 - 3.40 (3.67 – 3.40)	156-91- 3.53 (3.87 - 3.53)
Data processing						
Total reflections	400454 (34541)	335798 (34968)	476523 (24279)	211854(25842)	120281 (24903)	45676 (10425)
Unique reflections	15954 (1360)	13708 (1307)	23634 (1197)	12717(1460)	5025 (989)	4716 (1077)
Multiplicity	25.1 (25.4)	24.5 (26.8)	20.2 (20.3)	16.7 (17.7)	23.9 (25.2)	9.7 (9.7)
Completeness (%)	100.0 (100.0)	99.1 (98.5)	100.0 (100.0)	96.1 (100.0)	100.0 (100.0)	99.5 (99.1)
Mean <i>I</i> / σ (<i>I</i>)	43.6 (6.9)	32.1 (11.1)	26.8 (6.1)	28.2 (3.4)	22.2 (8.5)	9.7 (2.4)
<i>R</i> _{merge} ^a (%)	5.8 (59.8)	7.0 (38.5)	8.1 (60.1)	4.1 (63.6)	15.7 (57.8)	8.1 (43.5)
<i>R</i> _{rimb} ^b (%)	1.2 (12.2)	1.5 (7.6)	1.8 (13.6)	10 (15.4)	3.3 (11.6)	2.7 (14.6)
CC1/2	1.00 (0.98)	0.99 (0.99)	0.99 (0.95)	1.00 (0.97)	0.99 (0.99)	0.99 (0.99)
Asymetric unit	1	1	1	1	1	1
Wilson B factor (Å ²)	39.07	41.06	14.92	50.41	65.80	98.53

4

Supplementary Table S2: (Cont).

Crystal data	Nrd1 301-488 SeMet	Nrd1 301-488	Nrd1 290-468	Nrd1 290-468	Nrd1 290-468	Nrd1 290-468
Refinement						
<i>R</i> _{work} / <i>R</i> _{free} ^c (%)		20.68/23.30	17.76/20.80	19.67/21.72	18.11/21.59	22.51/26.86
N° of atoms/Mean B factors (Å²)		1373/59.95	1587/21.67	1448/62.97	1437/89.51	1447/128.46
Protein		1327/60.00	1389/20.20	1310/61.71	1310/85.70	1303/125.05
RNA		-	-	82/75.74	125/129.08	136/159.90
Ligand		4/74.00	60/28.66	16/107.47	-	8/169.30
Water molecules		42/57.70	138/33.22	38/54.51	1/43.95	2/69.17
Ramachandran plot (%) Favoured		99.0 / 0	99.0 / 0	98.0 / 0	98.0 / 0	96.0 / 1
/ Outliers						
R.m.s. Bonds (Å)/Angles (°)		0.010/1.31	0.007/1.25	0.009/1.36	0.007/1.22	0.007/1.16
PDB codes		5O1W	5O1X	5O1Y	5O1Z	5O2O
Missing residues in pdb		301-305/472-485	290-301/464-468	290-301/465-468	290-301/465-468	290-301/464-486

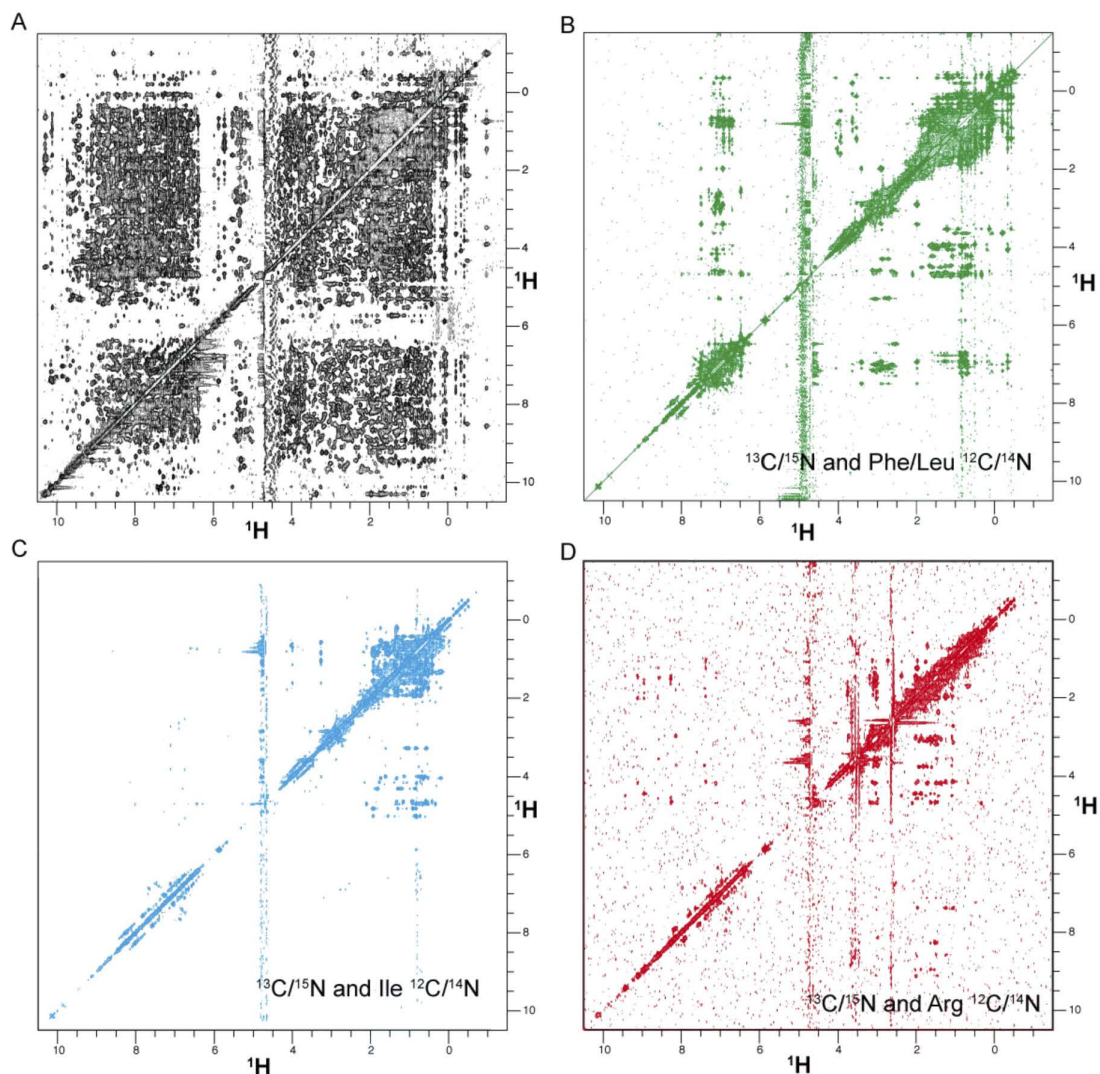
5

Supplementary Table S3. Summary of NMR restraints and structural calculation statistics for Nrd1₂₉₀₋₄₆₈ solution structure.

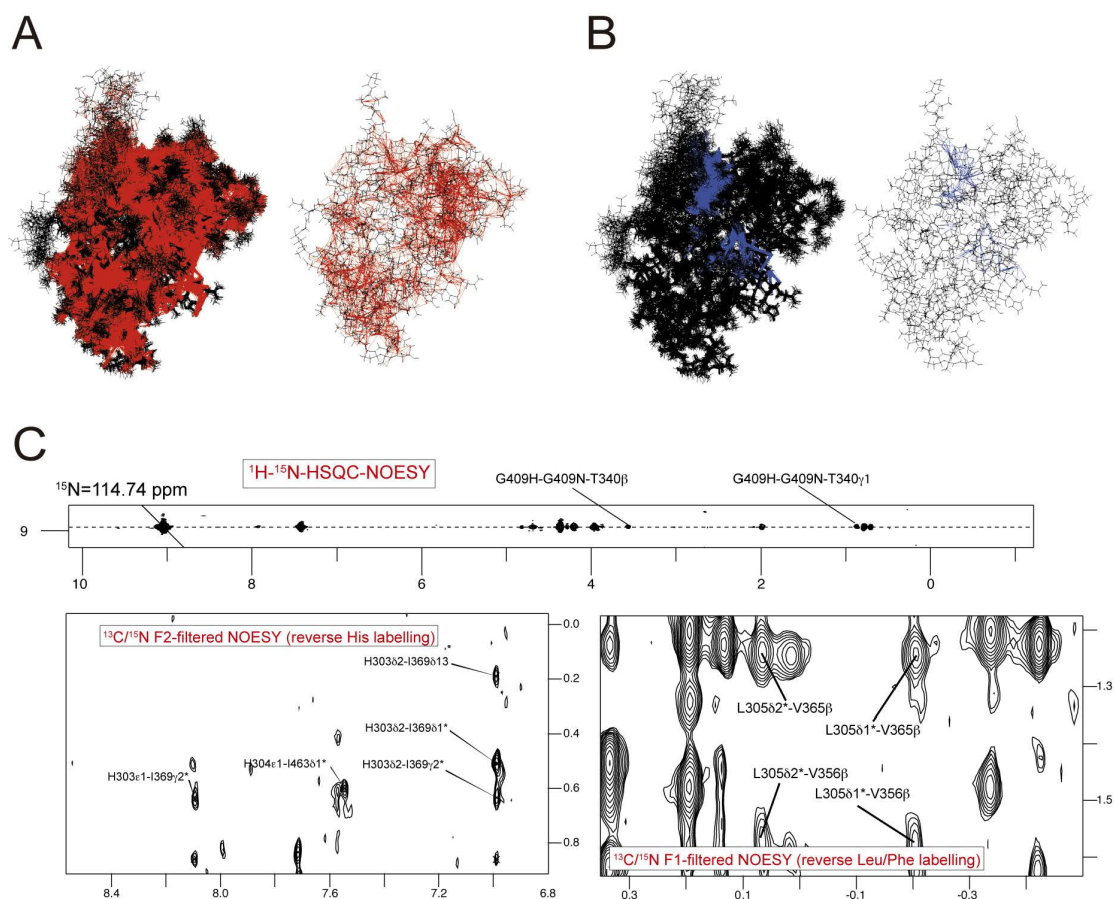
NMR experimental restraints	
<i>NOE-derived</i>	
Intraresidue	523
Sequential	756
Medium-range ($1 < i-j < 4$)	405
Long-range ($i-j > 4$)	1055
Total per residue	5.9
Intra RRM	1301
Intra SD	1284
SD to RRM	153
<i>TALOS+ obtained restraints</i>	
ϕ angle restraints	123
ψ angle restraints	131
Structure statistics	
<i>Mean AMBER energies (kcal/mol \pm SD)</i>	
Total	-4338 \pm 43
Van der Waals	-1380 \pm 20
Restraints (distance + angle)	6 \pm 1
<i>Violations</i>	
Distance*	5.3 \pm 1.3
Maximum distance violation (Å)	0.25
Angle**	4.2 \pm 0.6
Maximum angle violation (°)	18.7
<i>RMSD from ideal geometry</i>	
Bond lengths (Å)	0.010
Bond angles (°)	1.98
<i>Ramachandran Plot analysis</i>	
Residues in most favoured regions	95 \pm 1 %
Residues in additionally allowed regions	5 \pm 1 %
Residues in disallowed regions	0 \pm 0 %
<i>Averages RMSD to mean structure (range)</i>	
N, CO, C \checkmark (Å) (\pm SD)	0.63 \pm 0.11
All heavy (Å) (\pm SD)	0.97 \pm 0.11
*Averaged value per structure of distance violations > 0.20 Å \pm SD.	
** Averaged value per structure of total angle violations \pm SD.	

Supplementary Table S4. NOE-derived distance restraints between RRM and SD domains. The X-ray column shows NMR restraints violations > 0.4 Å found in the x-ray Nrd1₂₉₀₋₄₆₈ structure and reflect minor differences between both structures. Restrains for residues without coordinates in the X-ray are labelled as NA (not available).

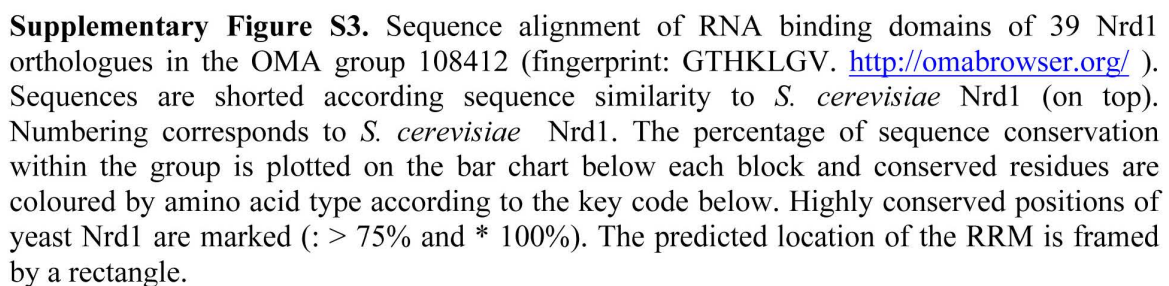
NMR upper limit restraint					X-ray	NMR upper limit restraint					X-ray
294 PRO HA	353 TRP HH2	5.71	NA			305 LEU QGD	352 GLU QB	4.51			
294 PRO HA	353 TRP HZ2	4.77	NA			305 LEU QGD	353 TRP HA	5.06			
294 PRO HB2	353 TRP HH2	4.76	NA			305 LEU QGD	353 TRP HD1	5.36			
294 PRO HB2	353 TRP HZ2	4.66	NA			305 LEU QGD	353 TRP HE1	5.08			
294 PRO HB3	353 TRP HH2	4.76	NA			305 LEU QGD	353 TRP HE3	4.44	0.8		
294 PRO HB3	353 TRP HZ2	4.66	NA			305 LEU QGD	353 TRP HH2	4.95			
294 PRO QB	353 TRP HH2	3.96	NA			305 LEU QGD	353 TRP HZ2	4.34			
294 PRO QB	353 TRP HZ2	3.99	NA			305 LEU QGD	353 TRP HZ3	3.98	1.2		
294 PRO QG	353 TRP HZ2	5.09	NA			305 LEU QGD	355 LEU QB	5.98	1.2		
296 THR HA	353 TRP HE1	4.55	NA			305 LEU QGD	356 ALA H	4.48			
296 THR QG2	353 TRP HE1	5.27	NA			305 LEU QGD	356 ALA QB	3.38	1.1		
296 THR QG2	353 TRP HZ2	4.45	NA			305 LEU QGD	365 VAL HB	4.31	0.7		
298 PHE HZ	352 GLU H	5.99	NA			305 LEU QGD	365 VAL QQG	3.67	2.0		
298 PHE HZ	352 GLU QB	4.68	NA			305 LEU QGD	367 SER H	4.83			
298 PHE HZ	353 TRP HD1	5.06	NA			305 LEU QGD	367 SER HA	4.41			
298 PHE HZ	353 TRP HE1	4.50	NA			305 LEU QGD	368 VAL HB	4.82			
298 PHE QD	353 TRP HE1	4.72	NA			305 LEU QGD	368 VAL QQG	3.44	1.7		
298 PHE QD	368 VAL QQG	5.36	NA			306 TYR HA	353 TRP HZ2	5.02			
298 PHE QE	352 GLU QB	3.85	NA			306 TYR QD	366 GLN HA	5.26			
298 PHE QE	352 GLU QG	4.12	NA			306 TYR QE	366 GLN HA	5.20			
298 PHE QE	353 TRP HD1	4.84	NA			317 HIS HD2	364 GLU HB2	4.38			
298 PHE QE	353 TRP HE1	4.57	NA			317 HIS HD2	364 GLU HB3	4.38			
298 PHE QE	353 TRP HZ2	4.54	NA			317 HIS HD2	364 GLU QG	4.16			
298 PHE QE	368 VAL QQG	5.29	NA			337 TYR QD	384 ARG QD	4.60			
303 HIS HA	368 VAL H	3.97				338 SER H	384 ARG HH11	4.72			
303 HIS HB2	368 VAL H	4.46				338 SER H	384 ARG HH12	4.72			
303 HIS HB3	368 VAL H	4.46				338 SER H	384 ARG QH1	4.10			
303 HIS HD2	369 ILE H	6.50				338 SER H	406 TRP HH2	5.34			
303 HIS HD2	369 ILE QD1	4.40	1.8			338 SER HA	340 THR H	4.55			
303 HIS HD2	369 ILE QG1	4.50				338 SER HA	406 TRP HE3	5.67			
303 HIS HE1	369 ILE HA	5.49				338 SER HB2	339 ARG H	5.88			
303 HIS HE1	369 ILE HG12	5.65				338 SER HB2	340 THR H	5.10			
303 HIS HE1	369 ILE HG13	5.65				338 SER HB3	339 ARG H	5.88			
303 HIS HE1	369 ILE QD1	5.01				338 SER HB3	340 THR H	5.10			
303 HIS HE1	369 ILE QG1	4.97				338 SER QB	340 THR H	4.29			
303 HIS HE1	369 ILE QG2	5.00				338 SER QB	406 TRP HE3	5.19			
303 HIS QB	368 VAL H	3.86				339 ARG H	459 GLU H	5.34			
303 HIS QB	369 ILE QD1	4.44				339 ARG H	459 GLU HA	5.61			
303 HIS QB	369 ILE QG1	4.59				339 ARG H	459 GLU QB	4.62			
304 HIS HA	368 VAL H	5.30				339 ARG H	459 GLU QG	6.27			
304 HIS HD2	367 SER HB2	4.93				339 ARG QH2	460 PRO HA	5.26			
304 HIS HD2	367 SER HB3	4.93				340 THR H	459 GLU HB2	6.10			
304 HIS HD2	367 SER QB	4.27				340 THR H	459 GLU HB3	6.10			
305 LEU H	367 SER H	4.14				340 THR H	459 GLU QB	5.22			
305 LEU H	367 SER HA	4.75				340 THR HB	408 VAL H	5.09			
305 LEU H	367 SER QB	5.25				340 THR HB	409 GLY H	4.96			
305 LEU HA	353 TRP HE1	6.50				340 THR QG2	407 GLY H	5.19			
305 LEU HA	353 TRP HZ2	4.62				340 THR QG2	408 VAL H	4.65			
305 LEU HG	353 TRP HZ2	4.34	0.9			340 THR QG2	409 GLY H	5.30			
305 LEU QD1	352 GLU HA	5.71	0.8			341 LEU QGD	407 GLY H	4.78			
305 LEU QD1	353 TRP HD1	6.50				378 PHE QD	407 GLY QA	4.66			
305 LEU QD1	353 TRP HE1	6.50				378 PHE QD	408 VAL H	5.24			
305 LEU QD1	353 TRP HE3	5.10				378 PHE QE	407 GLY QA	5.14	1.1		
305 LEU QD1	353 TRP HZ2	5.17				378 PHE QE	408 VAL H	5.01	1.4		
305 LEU QD1	353 TRP HZ3	4.73				378 PHE QE	408 VAL QQG	5.44	3.3		
305 LEU QD1	356 ALA H	5.45				378 PHE QE	418 TYR QE	5.24	3.6		
305 LEU QD1	356 ALA QB	4.22				382 TYR HA	443 TRP HD1	4.46			
305 LEU QD1	365 VAL HB	4.96				382 TYR HA	443 TRP HE1	5.89			
305 LEU QD1	367 SER H	5.50				382 TYR QB	443 TRP HD1	4.74			
305 LEU QD1	367 SER HA	5.26				382 TYR QB	443 TRP HE1	5.10			
305 LEU QD1	368 VAL H	5.10				406 TRP H	407 GLY H	5.34			
305 LEU QD1	368 VAL QQG	4.01	1.4			406 TRP H	407 GLY QA	5.21			
305 LEU QD2	352 GLU HA	5.71				406 TRP H	418 TYR HA	5.84			
305 LEU QD2	353 TRP HD1	6.50				406 TRP HD1	418 TYR HA	4.94			
305 LEU QD2	353 TRP HE1	6.50				406 TRP HD1	419 GLN HA	4.50			
305 LEU QD2	353 TRP HE3	5.10	1.0			406 TRP HD1	421 GLY HA2	4.42			
305 LEU QD2	353 TRP HZ2	5.17				406 TRP HD1	421 GLY HA3	4.42			
305 LEU QD2	353 TRP HZ3	4.73	1.6			406 TRP HD1	421 GLY QA	3.84			
305 LEU QD2	356 ALA H	5.45				406 TRP HE1	419 GLN HA	6.36			
305 LEU QD2	356 ALA QB	4.22	1.5			406 TRP HE1	420 HIS HA	4.59			
305 LEU QD2	365 VAL HB	4.96	1.6			406 TRP HE1	421 GLY H	6.30			
305 LEU QD2	367 SER H	5.50				406 TRP HE1	421 GLY QA	5.47			
305 LEU QD2	367 SER HA	5.26				406 TRP HE3	407 GLY H	4.06			
305 LEU QD2	368 VAL H	5.10				406 TRP HE3	421 GLY HA2	5.67			
305 LEU QD2	368 VAL QQG	4.01	1.8			406 TRP HE3	421 GLY HA3	5.67			
305 LEU QGD	352 GLU HA	4.91	1.1			406 TRP HE3	421 GLY QA	4.95			
						406 TRP QB	407 GLY H	4.35			

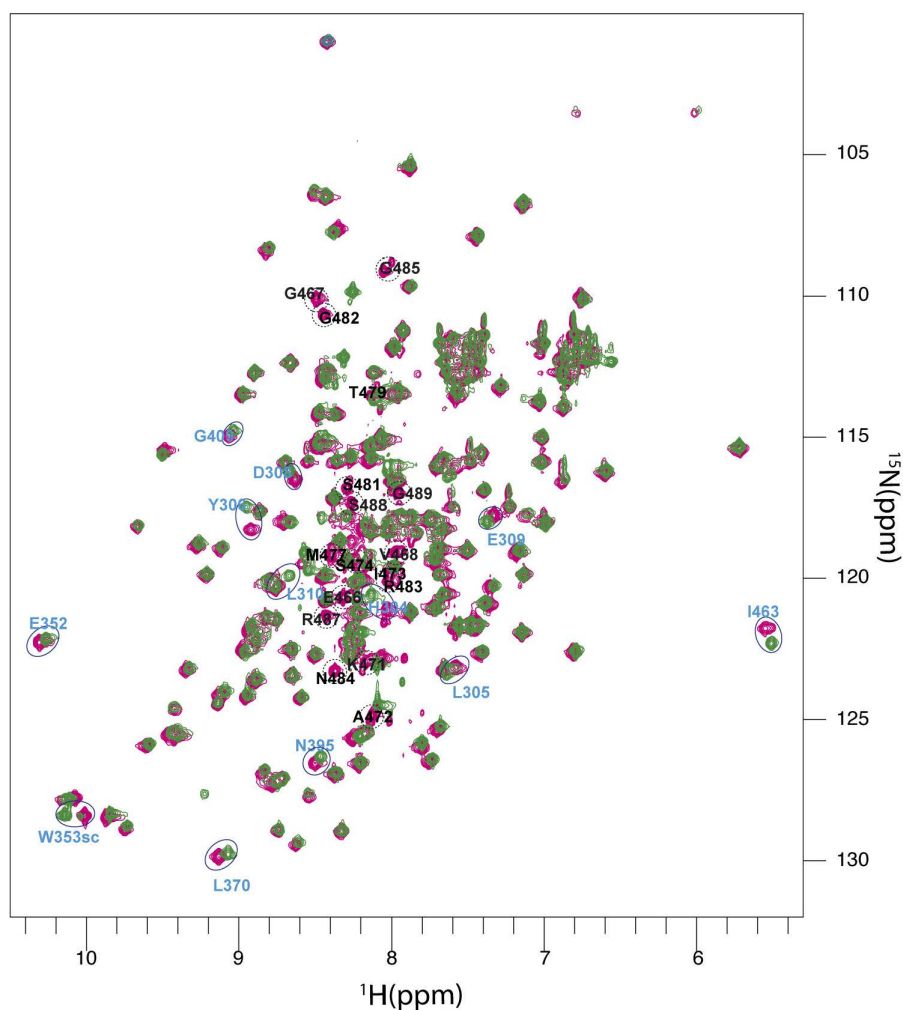


Supplementary Figure S1. Selectively unlabelled samples in $^{13}\text{C}/^{15}\text{N}$ background greatly facilitates NOEs assignments. Comparison between the standard 2D NOESY (A) and 2D $^{13}\text{C}/^{15}\text{N}$ double-filter NOESY of three samples obtained by selectively unlabelling Phe/Leu (B), Ile (C) and Arg (C). Samples were produce by growing bacterial cultures in $^{13}\text{C}/^{15}\text{N}$ medium and adding the aminoacid/metabolic precursor 30 min prior induction. Metabolic precursors used for each amino acid (Sigma-Aldrich) are explained in detail in (4,5) which the exception of L-citrulline and L-histidinol (Sigma-Aldrich) that have been used in this work for the first time for selective unlabelling Arg and His respectively.

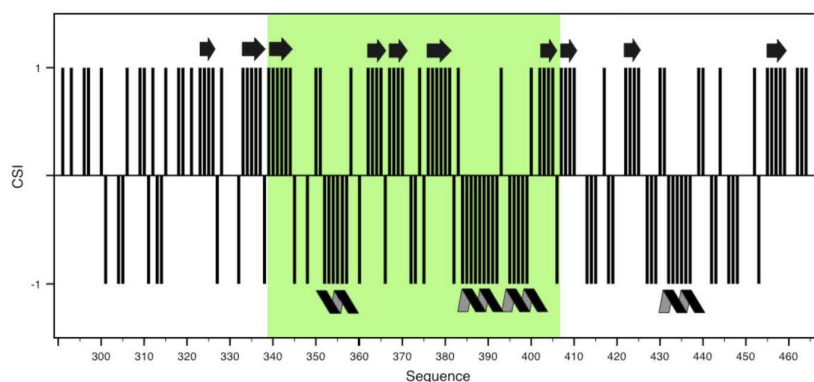


Supplementary Figure S2. Selected NMR data and derived distance restraints used in the calculation of 5O1T (full restraint set available at the PDB site). **A** Representation of all distance restraints (in red) over the NMR ensemble/one conformer (in black) r. **B** Representation of SD-RRM restraints subset (blue). **C** Zooms in of selected regions of various types of NOESY spectra. For simplicity, only the SD to RRM NOEs have been labelled.

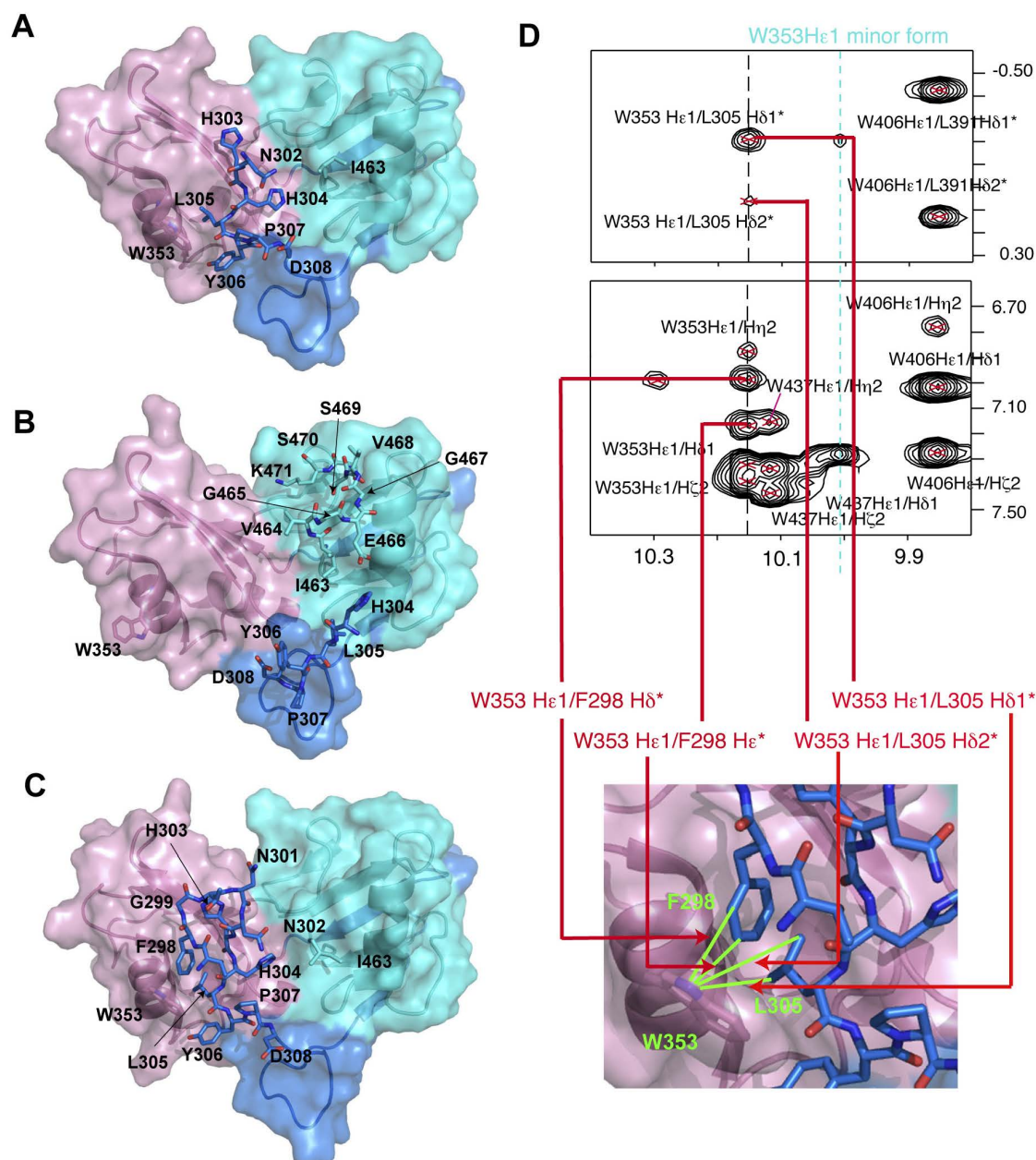




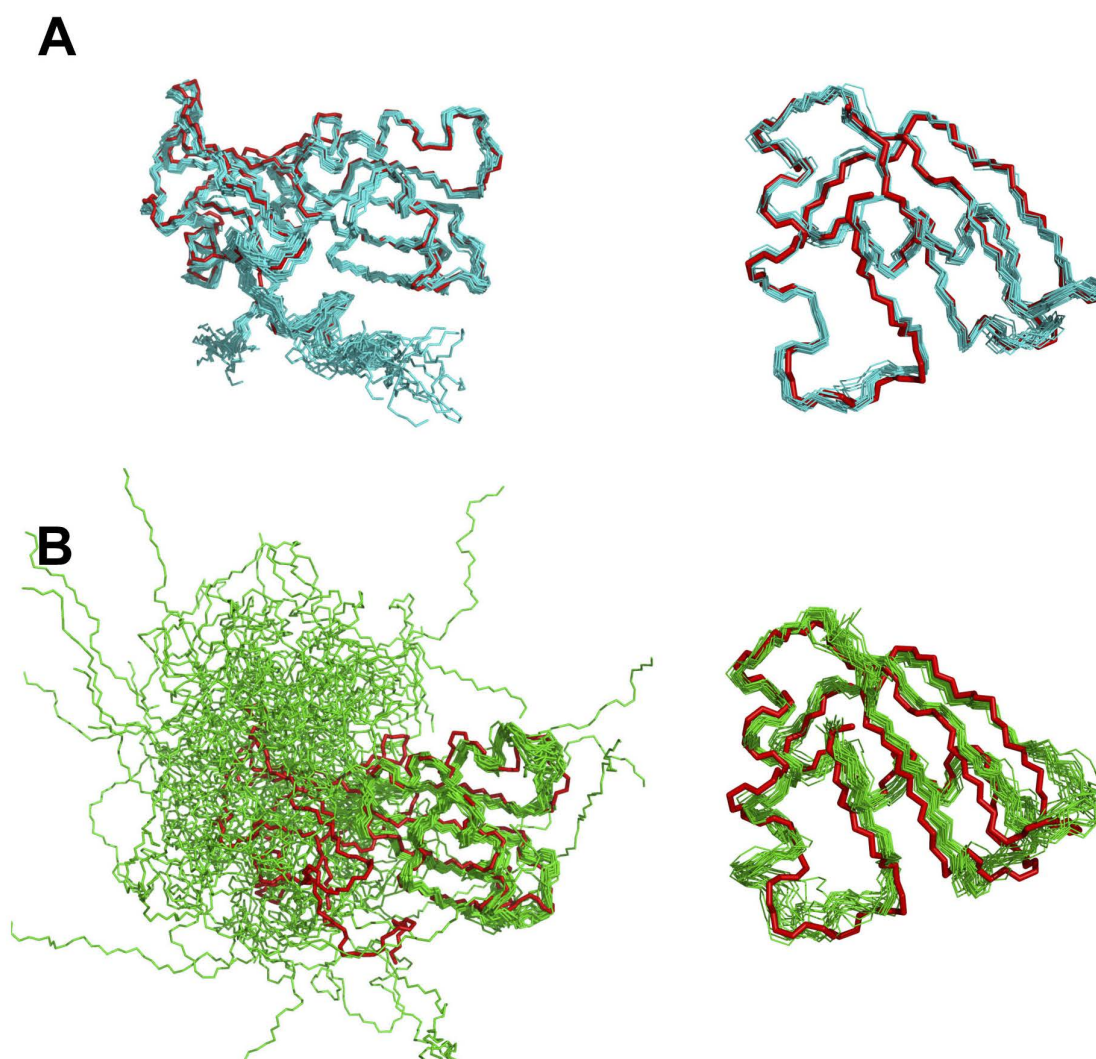
Supplementary Figure S4. Comparison of ^1H - ^{15}N HSQC NMR spectra of Nrd1₃₀₁₋₄₈₉ (magenta) and Nrd1₂₉₀₋₄₆₈ (green). Signals for residues in the unfolded Nrd1₃₀₁₋₄₈₉ C-terminus (residues 466-489) have been marked (dashed circled and black labels). Crosspeaks exhibiting larger differences between the two constructs have been marked (solid ovals and cyan labels). Both spectra have been recorded at identical experimental conditions (see materials and methods) in a Bruker AV 800 MHz spectrometer.



Supplementary Figure S5. Chemical Shift Index by residue for Nrd1₂₉₀₋₄₆₈ RNA binding domain and predicted secondary structure elements (1: β -sheet, -1: α -helix). CSI defined from $^{13}\text{C}\alpha$, $^{13}\text{C}'$ and $^{13}\text{C}\beta$ chemical shifts. RRM predicted limits are shadowed in green.

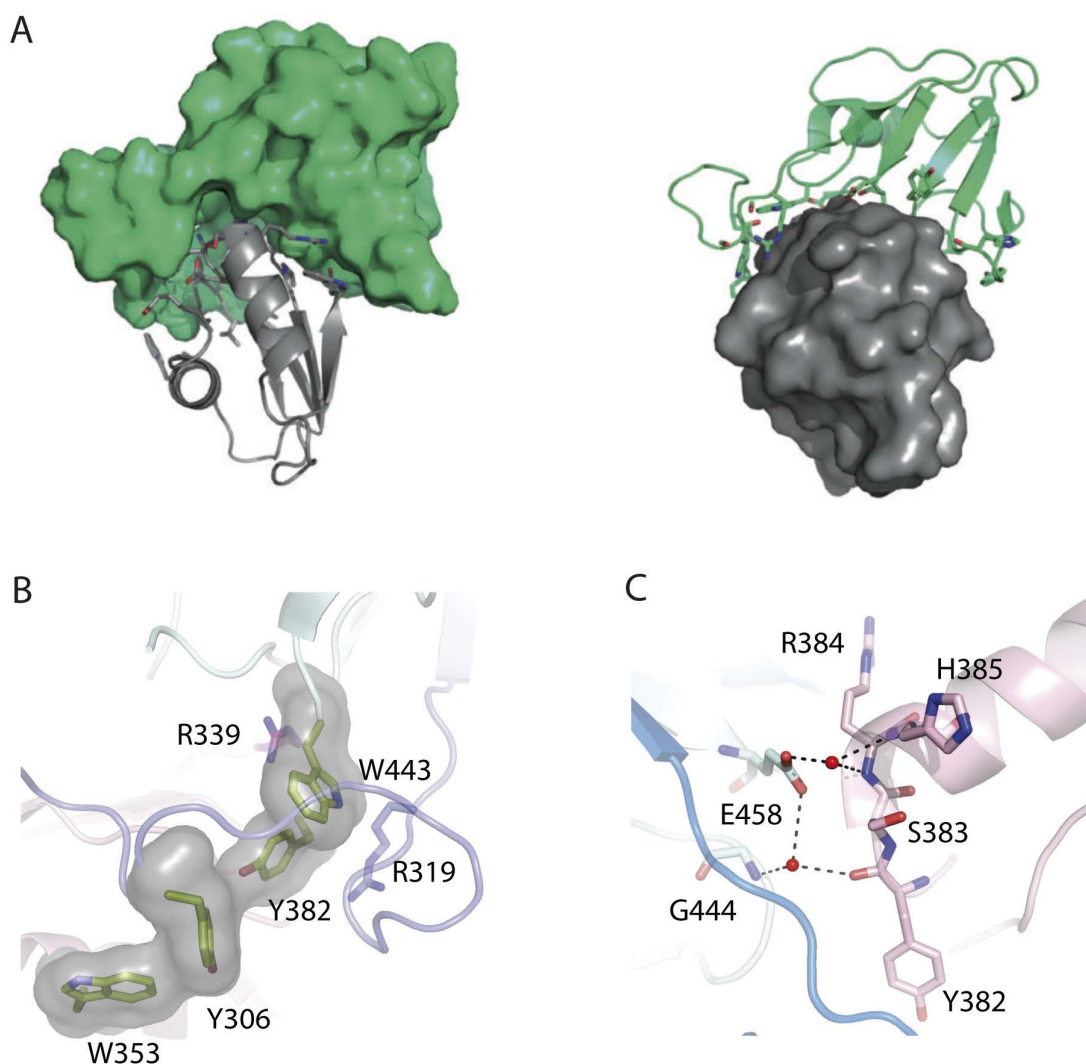


Supplementary Figure S6. Nrd1RBD structures show conformational differences at their termini. **A** Nrd1₂₉₀₋₄₆₈ crystal structure. **B** Nrd1₃₀₁₋₄₈₉ crystal structure. **C** Nrd1₂₉₀₋₄₆₈ NMR structure (a representative conformer). The unstructured 290-297 segment omitted for clarity. Domains coloured with the same code as in Figure 1. The Nrd1 RBD structural core (residues 308-462) has been represented as ribbon and translucent surface. The C- and N-termini of different length in the two Nrd1 RBD constructs have been represented in sticks and each residue labelled. Trp 353 of the RRM has been also labelled represented in sticks to show the differences on its conformation among the structures. **D.** Detail of 2D NOESY showing the NOEs involving various Nrd1 tryptophan H ϵ 1 protons. Long-range W353 NOEs are labelled in red and correlated with their corresponding distance restraint in the structure (zoom in of C). The position of W353 H ϵ 1 minor form is represented by a dashed cyan line.

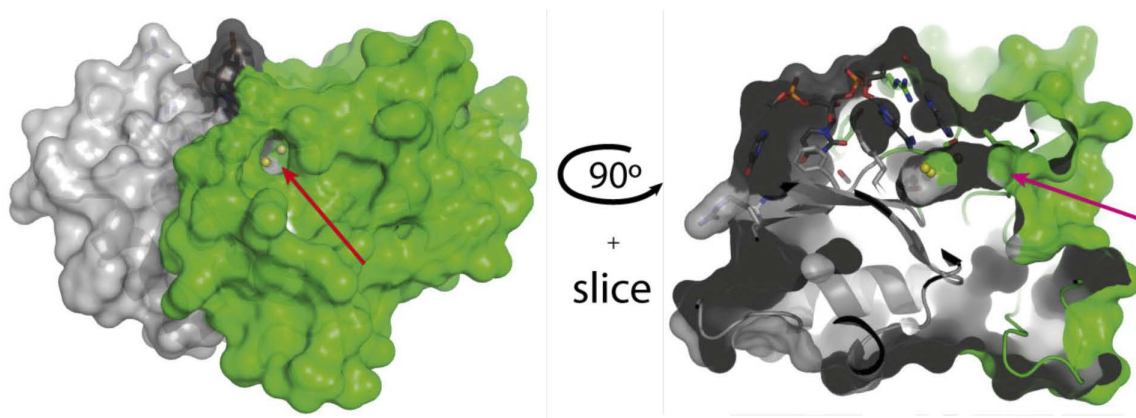


Supplementary Figure S7. Comparison between X-ray and NMR structures of Nrd1 RBD.

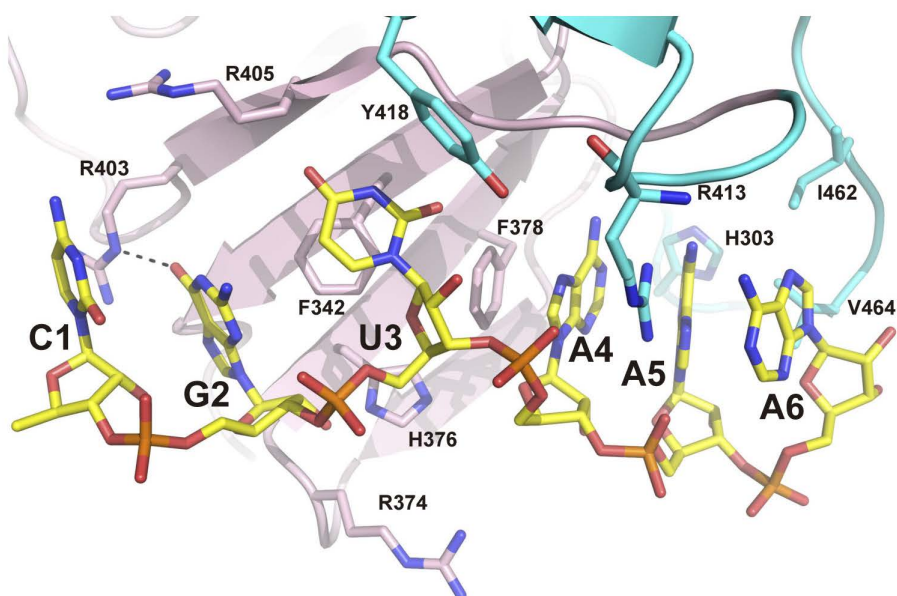
A. Superposition between the backbone of X-ray (red) and NMR (cyan) structures of Nrd1₂₉₀₋₄₆₈ (this work). Detail of the RRM alignment is shown on the right. **B.** Superposition between the backbone of Nrd1₂₉₀₋₄₆₈ X-ray structure (in red) (this work) and previously determined NMR of Nrd1₃₀₇₋₄₉₀ (PDB:2m88) (in green) (6), including a comparison of the RRMs domains on the right. The lack of definition of the SD domain in the Nrd1₃₀₇₋₄₉₀ might be due to partial truncation of the N-terminus of this subdomain or, most likely to protein misfolding, caused by overexpression at 30°C. Our earlier studies showed evidences that our construct Nrd1₃₀₁₋₄₈₉ is misfolded when expressed at temperatures of >12°C: inhomogeneous signals in the ¹H-¹⁵N-HSQC, faster degradation and unsuccessful crystallization



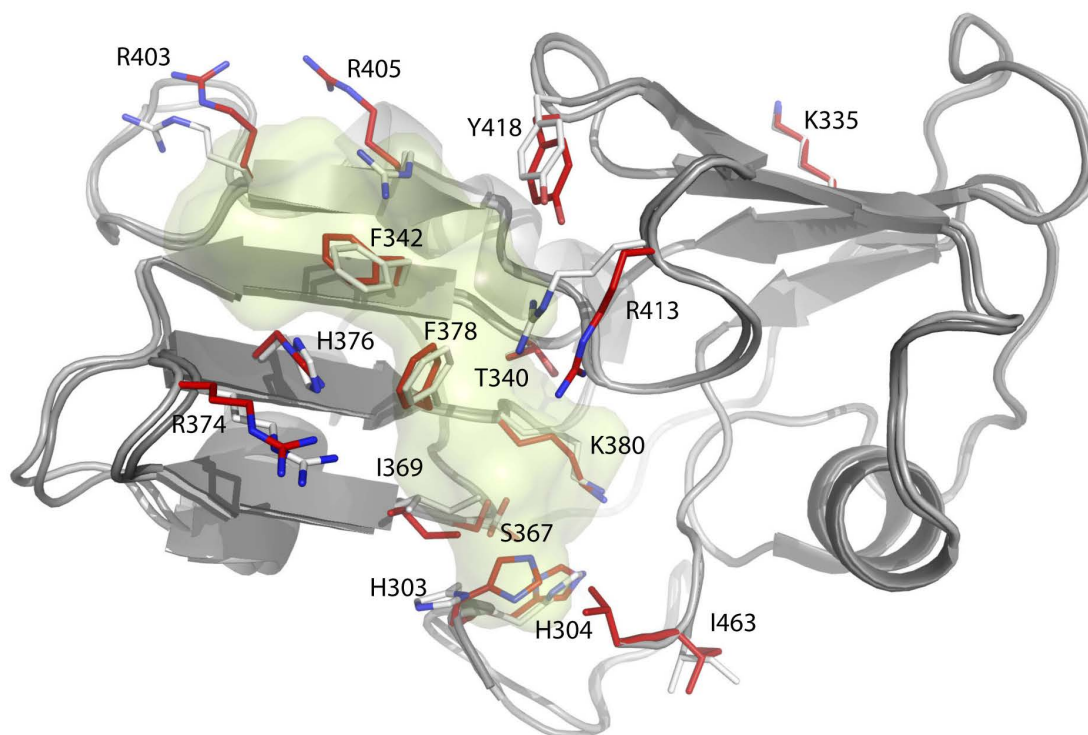
Supplementary Figure S8. Nrd1 RRM and SD subdomains interface details. **A.** RRM (grey) and SD (green) interact through a curved interface. **B.** Aromatic layer separating the Arg319 and Arg339 polar clusters. The residues forming the layer are shown as green sticks and surrounded by a grey surface representation. **C.** Only two structural waters are found at the SD-RRM interface despite its polar nature. The two water molecules (shown as red spheres) make hydrogen bonds with residues from both domains and stabilise the buried Glu458 residue. Equivalent waters are built in crystal structures with best resolution from constructs Nrd1₂₉₀₋₄₆₈ and Nrd1₃₀₁₋₄₈₉.



Supplementary Figure S9. The structure of the Nrd1-RNA complex contains an internal water pocket (4 molecules; yellow). Magenta arrows show the entry access to this pocket.



Supplementary Figure S10. Structure of Nrd1₂₉₀₋₄₆₈ in complex with CGUAAA obtained in phosphate buffer. Colour code is as in Figure 3. The structure reveals the hydrogen bond Arg403 N ϵ -Gua2 O6 that was not detected in the complex with GUAA (Figure 3) due to the interference of a tris molecule (present in the protein buffer in that case).



Supplementary Figure S11. The Nrd1 RBD experiences minor changes upon GUAA recognition. Superposition of the crystal structures of Nrd1₂₉₀₋₄₆₈ in its free (sidechain carbons coloured in white), and GUAA-bound (sidechains in red) states. The location of the GUAA sequence in the complex is shown as green transparent surface. The residues in contact with the RNA in the complex have been represented. The rest of the residues do not suffer significant changes upon binding. Arg403 and Arg405 are poorly defined in the structure of the free form and interact with a tris molecule in the complex (not shown here), thus their changes are probably not attributed to RNA recognition. The side chain of Lys335, mentioned several times in the discussion, has been included to emphasize its opposite position to the RNA interface.

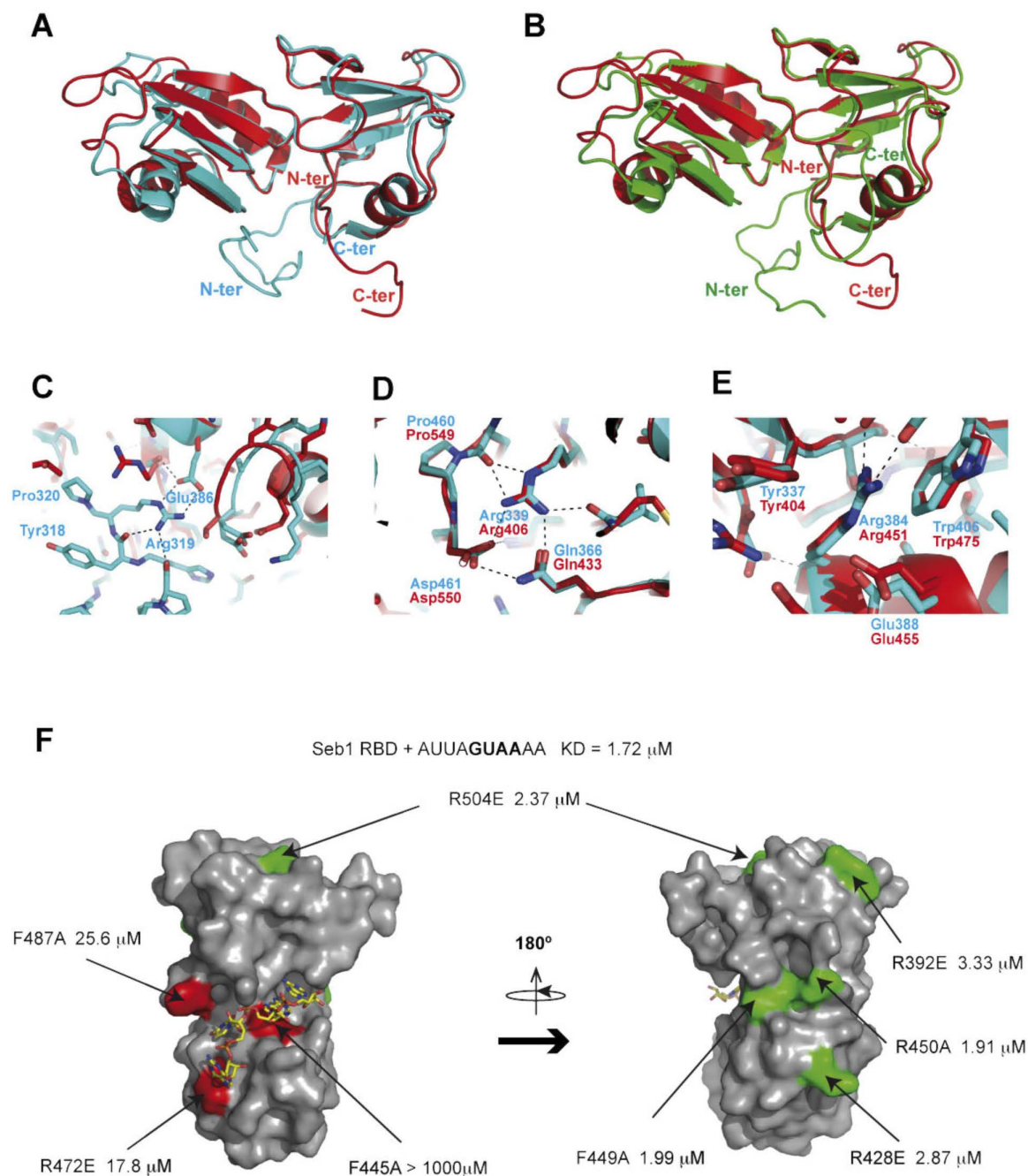
```

S. cere.      DSGRNGPRSGKPNKSGSISS-----ISPVP-YGNAPLASPPP-----QQYVQPM 522
T. pha.      DSIRSQQNYQNMVHHHP-----QPPQQQQQ--MMYGGYQ-----D-----TKYNNNN 544
K. pol.      DSNRNGPRSGKGEMVPTQ-----NVPQYGAYP-----PE-----Q----- 538
S. cas.      DS-RHGPKSAKAFNSPPAAA-----PIPMQQTQGMYPYST-N-----QMYPQ-- 523
S. dai.      DGTRNGPRSLKRNPMPLQSGSYATPPQQQPP--QMVQPQPYGSPN-----QLYGQLQ 566
C. gla.      DSGRNGPRSNKAKNMAPQYGP-----PQAAVISPSQGYAGMQQGYPQQQQQQMGMYGQP 550
C. col.      DSGRNGPRSGKSTKTSV--S-----SPSMPGYATNQIASPG-----QMYSQ-- 529
Z. rou.      DSGRNGPRSGKNSRSLSV-----SPPVTKYSMAPVPSPG-----HVGQ-- 517
* . * . :

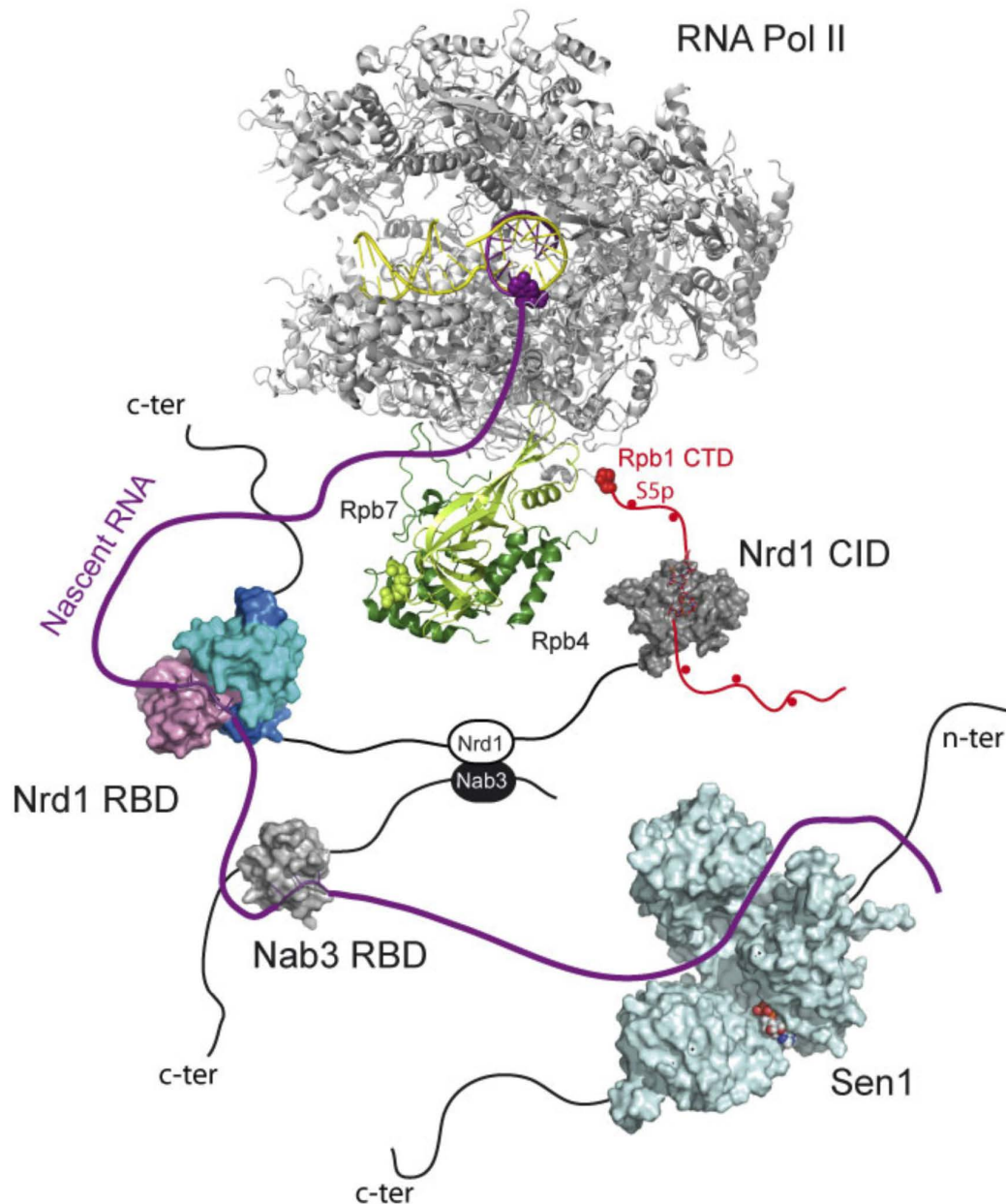
MQQPY-----GYAPNQ-----PLPSQ--GPAAAPVPVPPQQQFDPTAQLNSLMNMLNQQQQQQQS 575
FMPPPPPPQQYQNMHPPQPHMP-----HMP-----PQPHMPPQQQPQGFDPDPTAQLNSLMNMLNQQPPR---- 602
RYSGYPPQQNMGYNQPMPPQGVDE---YQPPPPPLPPQ-----QQQQPQQSFDPTAHLNSLMNMLNQQQR---- 601
MQGPPPPPPQQMGYQ--P-----YPQPQQAHPAPPPPPQQQGNIDPTAQLNSLMNMLNQQQQQRN--- 579
--PNMLPPPPQQQQSYAMN-----GQQQVPPPPP-----PQQQVQQNVADPTAQLNSLMNMLNQQQQRS--- 624
MQQPINAYAS-----PAQQQ-----PIMQAPPPAP-----PSQQQSSNYDPTAQLNSLMNMLNQQQQ--- 604
MQQPYPPQQAAPQQAAPPYGP-----PVAAPM-A-GYTQPPQQQPFDPDPTAQLNSLMNMLNQQQN----- 588
IPPAYPQPVAAYNGTPIAAAPQMAPAAATPSYPPAPAPQ--QAPAPPVPQQQSFDPDPTAQLNSLMNMLNQQQT--- 590
* :*:*****.:*.*

```

Supplementary Figure S12. Conservation of C-terminal elements of Nrd1 in different yeasts: *Tetrapispora phaffii*, *Kluyveromyces polysporus*, *Saccharomyces castellii*, *Saccharomyces dairenensis*, *Candida glabrata*, *Candida colliculosa* and *Zygosaccharomyces rouxii*



Supplementary figure S13. Structural comparison between Nrd1 and Seb1 (PDB: 5MD4) RNA binding domains. **A.** Superposition between Nrd1₂₉₀₋₄₆₈ (cyan) and Seb1 (red) crystal. **B.** Superposition between Nrd1₃₀₁₋₄₈₉ (green) and Seb1 (red) crystal. **C-E.** Detailed view of the three arginine-centred clusters at the Nrd1₂₉₀₋₄₈₆ RRM-SD interface (Figure 1E-G) superimposed with the equivalent regions of Seb1. Key residues in Nrd1 and Seb1 have been labelled in cyan and red respectively. **F.** Structure model of GUAA recognition by Seb1. The RNA coordinates have been taken from the Nrd1₂₉₀₋₄₆₈:GUAA complex after structural alignment between Nrd1 and Seb1 RBDs. Seb1 has been represented in grey surface and mutation sites (7) has been coloured according to their effect on RNA affinity (in red $K_D > 10$ -fold of wt; in green $K_D < 3$ -fold of wt). Mutants have been labelled together with their corresponding K_D . The K_D of wild-type Seb1 and the RNA probe used in the mutagenesis experiments are presented above for reference.



Supplementary figure S14: Structural model of the NNS complex and elongating RNA polymerase. RNA polymerase II is shown in cartoons with DNA coloured in yellow, nascent RNA in purple and the stalk subunits: Rpb7 and Rpb4 in two shades of green. Nascent RNA has been represented as a purple line showing the interactions with Rpb7 and Sen1, with the exception of the regions of GUAA and UCUU terminators bound to Nrd1 (this work) and Nab3 (PDB: 2XNR) RBD. The Rpb7 region that interacts with Nrd1 is shown as green spheres (8). The Rpb1 CTD has been represented as a red line connecting the last Rpb1 residue seen in the RNA Pol II structure (1Y1W) and the CTD Ser2p peptide on the Nrd1 CID-CTD complex (PDB: 1LO6). The structure of Sen1 (PDB: 5MZN) is on its RNA-free state; the expected location of the RNA has been drawn in an equivalent position to that reported for Upf1 RNA helicase complex (PDB: 2XAO not shown in this picture). All the atomic models have been represented at the same scale. The loading order of the Nrd1, Nab3 and Sen1 on the nascent RNA does not reflect the biochemical knowledge about the real order in the NNS complex.

REFERENCES

1. Steinmetz, E.J. and Brow, D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.*, **16**, 6993-7003.
2. Steinmetz, E.J., Ng, S.B.H., Cloute, J.P. and Brow, D.A. (2006) cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Mol. Cell. Biol.*, **26**, 2688-2696.
3. Vasiljeva, L. and Buratowski, S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239-248.
4. Prasanna, C., Dubey, A. and Atreya, H.S. (2015) Amino acid selective unlabeled in protein NMR spectroscopy. *Methods Enzymol.*, **565**, 167-189.
5. Rasia, R.M., Brutscher, B. and Plevin, M.J. (2012) Selective isotopic unlabeled of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. *ChemBioChem*, **13**, 732-739.
6. Bacikova, V., Pasulka, J., Kubicek, K. and Stefl, R. (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.*, **42**, 8024-8038.
7. Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.H., Kecman, T. *et al.* (2017) The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat. Commun.*, **8**, 14861.
8. Mitsuzawa, H., Kanda, E. and Ishihama, A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.*, **31**, 4696-4701.

3.2.1 Discusión trabajo de investigación 3

El trabajo estructural sobre Nrd1 proporciona información útil para comprender mejor las bases moleculares del funcionamiento del complejo NNS, a la vez que nos permite avanzar en la comprensión del reconocimiento de ARN por proteínas, y en concreto por los dominios RRM.

Con el fin de obtener cristales del dominio de unión a ARN de Nrd1, el grupo del Dr. Pérez-Cañadillas diseñó y produjo una serie de construcciones basadas en un análisis de secuencias de los ortólogos de Nrd1. Este alineamiento reveló que, además de conservar un dominio RRM canónico (residuos 339-407), también conservan las regiones que lo flanquean, sugiriendo que el dominio completo de unión a ARN podría ser más grande de lo esperado. Se cristalizaron tres construcciones distintas de Nrd1 (Nrd1₃₀₁₋₄₈₉, Nrd1₃₀₁₋₄₆₈, Nrd1₂₉₀₋₄₆₈) basadas en este análisis y en datos de RMN. La estructura se resolvió a partir de los cristales de Nrd1₃₀₁₋₄₈₇ sustituida con Selenometionina (Se-Met) y la técnica de dispersión anómala simple (SAD). El análisis de todas ellas mediante difracción de rayos X demostró que todas las construcciones tienen una estructura similar, presentando solo algunas diferencias en el extremo N-terminal (antes del Asp308) y el C-terminal (después del residuo Ile463). El grupo del Dr. Pérez Cañadillas determinó la estructura de una de estas construcciones mediante RMN. En contraste con una estructura previa en disolución, la estructura global de Nrd1 en el cristal y en solución son muy similares mostrando únicamente ligeras diferencias en los primeros residuos de la construcción.

La estructura del dominio de unión a ARN de Nrd1 muestra un plegamiento inusual; el dominio RRM está fusionado con un segundo dominio denominado SD (dominio dividido o “*Split Domain*”) construido a partir de dos segmentos no consecutivos en secuencia: 304-338 y 408-464 (Figura 1A-B del Trabajo de Investigación 4). El SD también posee un plegamiento tipo α/β pero diferente al del RRM (ver Introducción, sección 1.5.1). Topológicamente, la estructura general del dominio de unión a ARN de Nrd1 podría describirse como una inserción del RRM en uno de los bucles del SD. Es bastante frecuente encontrar dominios RRM con características diferenciadas presentando regiones adicionales en los extremos C y/o N-terminales (Oberstrass et al., 2005) como hélices α (Jacks et al., 2003; Perez Canadillas y Varani, 2003) y hojas β (Martínez-Lumbreras et al., 2016). Sin embargo, en Nrd1 la configuración del dominio RRM es completamente novedosa. El único homólogo estructural para el dominio SD, reportado mientras nuestro trabajo estaba en vías de publicación, pertenece a la proteína Seb1 de *Schizosaccharomyces pombe*, homóloga de Nrd1 (Wittmann et al., 2017). Por otro lado, la falta de plegamiento en el dominio SD contenido en la estructura de Nrd1 publicada por el grupo del Dr. Stefl (Bacikova et al., 2014) puede deberse a que las construcciones y los protocolos de expresión y purificación elegidos por estos autores son muy diferentes.

En cuanto a la relación entre ambos dominios, RRM y SD, estos interaccionan a través de una interfaz polar de gran tamaño (Figura 1D-G del Trabajo de Investigación 4). Esta interfaz intramolecular es mantenida por una red de tres residuos de arginina: Arg319, Arg339 y Arg384. Cabe esperar que, en hongos, las proteínas homólogas conserven la estructura general incluida la interfaz RRM-SD, puesto que la mayoría de los residuos involucrados se encuentran totalmente conservados. Sin embargo, el dominio SD de Nrd1 incluye extensiones más largas en el N-terminal que el dominio equivalente en Seb1 lo cual se traduce en un mayor número de contactos con el dominio RRM. Esto, unido al hecho de que Seb1 no conserva el residuo Arg319, sugiere que la formación de la interfaz sea ligeramente distinta.

Uno de los principales objetivos del trabajo ha sido conocer los detalles del reconocimiento de ARN por Nrd1. Se sabe que el motivo GUAA se encuentra en la mayoría de los transcritos procesados a través de la vía NNS (Jamonnak et al., 2011; Schulz et al., 2013). Mediante calorimetría de titulación isotérmica (ITC) y ensayos de anisotropía de fluorescencia se ha determinado que Nrd1₂₉₀₋₄₆₈ reconoce GUAA con una afinidad superior a GUAG y, ambas con una afinidad significativamente más grande que las sondas AUAA/U y AUUA (ver Figura 2 y Tabla 1 del Trabajo de Investigación 4). La elevada afinidad en la interacción Nrd1: GUAA detectada (2 μ M) sugería que además de los residuos canónicos de los RRM podrían estar implicados más residuos en esta unión, como así fue corroborado en las estructuras de los complejos Nrd1-ARN (ver Figura 3 del Trabajo de Investigación 4). En todos los complejos que se obtuvieron, se utilizaron ARNs que contenía el motivo anterior (GUAA, CGUAAA y UUAGUAAUCC). Las bases adicionales al motivo no interaccionan de forma evidente con la proteína, lo cual, junto con el hecho de que el núcleo de GUAA se encuentre en la misma conformación en los tres complejos obtenidos, sugiere que Nrd1 reconoce principalmente dicho tetranucleótido. Por otra parte, al comparar la estructura libre y en complejo de Nrd1, se observa que la unión al ARN sólo causa ligeros cambios estructurales en Nrd1, lo que sugiere que el sitio de unión a ARN está estructuralmente preconfigurado. Las distintas estructuras del complejo revelaron que ambos dominios, RRM y SD, contribuyen al reconocimiento de GUAA, si bien el SD desempeña un papel de andamiaje proporcionando soporte a varias moléculas de agua que hacen el papel de puente entre el ARN y la proteína. En cuanto a la interacción del ARN con el RRM se produce en la clásica orientación 5' a 3', como en todos los complejos RRM-ARN que se conocen hasta la fecha. Una descripción completa de la interacción Nrd1-ARN está bien detallada en el trabajo de investigación (ver Figura 3 del Trabajo de Investigación 4). Nuestro trabajo permite extender las conclusiones a la proteína homóloga Seb1. El valor de la afinidad de Seb1 por AUUAGUAAAA (1,8 μ M) es muy cercano a la afinidad de Nrd1 por CCGUAACC (1,5-2,1 μ M) y ambas secuencias de ARNs contienen el módulo GUAA. Todas estas similitudes sugieren que Seb1 podría tener un modo de unión de ARN similar a Nrd1, donde el sitio de unión al ARN está también presumiblemente preformado. Es más, proponemos que este modo de unión estará conservado en toda la familia con pequeñas variaciones.

Estudios genéticos y funcionales llevados a cabo en este trabajo han puesto de manifiesto que la eliminación total o parcial del dominio SD es letal en la célula, por lo que este dominio es fundamental para la viabilidad celular (ver Figura 4A y 4C del Trabajo de Investigación 4). La singularidad de la estructura del dominio de unión a ARN de Nrd1 plantea la cuestión de si este dominio, además de participar en el reconocimiento de ARN de forma directa, posee alguna otra función. De hecho, se sabe que la región C-terminal de Nrd1 participa en redes de interacciones con otras proteínas además del complejo NNS. Así, Nrd1 interacciona con la subunidad Rpb7 de la ARN Pol II a través de su extremo C-terminal, y aunque se desconoce el mecanismo, estudios preliminares por RMN (Grupo Pérez-Cañadillas) indican que el dominio SD pudiera estar involucrado. Puesto que el heterodímero Rbp4/Rbp7 de la ARN Pol II participa en el reclutamiento del ARN, la interacción entre Nrd1 y Rbp7 podría ayudar a ubicar el dominio de unión a ARN de Nrd1 en la dirección del ARN recién sintetizado, permitiendo así un reconocimiento precoz de las secuencias GUAA/G que, en consecuencia, desencadenaría la terminación de la transcripción (ver Figura S14 del Trabajo de Investigación 4). Por otro lado, no se sabe si la función de Nrd1 opuesta a la señalización de RAS y relacionada con la supresión de genes tiene algo que ver con su reconocimiento de ARN o por el contrario con otros mecanismos diferentes. Aunque aún queda mucho por estudiar en la biología de Nrd1 y aclarar todos

sus mecanismos de actuación, este trabajo constituye un gran avance en el entendimiento de la función del complejo NNS en la terminación de la transcripción.

4. MATERIAL INÉDITO

4. Material inédito

4.1 Estudio estructural de la IP₆K

4.1.1 Introducción

Los IPPs o inositoles pirofosfatados son pequeñas moléculas solubles derivadas de inositol y ricas en energía, que contienen grupos monofosfato y uno o más grupos difosfato y están presentes en todas las células eucariotas, desde levaduras hasta mamíferos. Los IPPs son considerados mensajeros metabólicos y biosensores de energía, al coordinar el flujo energético con las vías de señalización. En consecuencia, muestran efectos pleiotrópicos en las células. Los IPPs son sintetizados a partir del IP₆, mediante dos familias de proteínas que actúan en diferentes posiciones del anillo: las IP₆Ks (familia de IPKs) y las PPIP5K (familia “ATP-*grasp*”). Las tres isoformas de IP₆K (IP₆K1, IP₆K2 e IP₆K3) han sido objeto de estudio en esta tesis. Además de su función catalítica en la síntesis de IPPs, las IP₆Ks interaccionan con otras proteínas, participando en múltiples procesos celulares como la exocitosis, la diabetes, la respuesta a la infección y la apoptosis.

Actualmente solo se conoce la estructura de la IP₆K de *Entamoeba histolytica* (*EhIP₆K*) (Wang et al., 2014a), un organismo protozoo parásito anaerobio muy primitivo. Mediante un análisis de secuencias se ha establecido que *EhIP₆K* posee claras diferencias con la isoforma humana (*hIP₆K*). Ambas comparten tan solo un 25% de identidad y un 41% de similitud, presentando la isoforma humana dos largas inserciones, una mayor de 100 residuos en el lóbulo N-terminal (inserto I) y otra de 45 residuos en el lóbulo C-terminal (inserto II). Estas diferencias hacen que la determinación de la estructura de la isoforma humana siga siendo objeto del máximo interés. La resolución estructural de esta proteína permitiría conocer su mecanismo de acción, profundizar en el conocimiento de las múltiples interacciones con otras proteínas, así como determinar la relación con el resto de las IPKs.

4.1.2 Objetivos

Los objetivos concretos en relación a la familia de las IP₆Ks se detallan a continuación:

- Expresión y optimización de las tres isoformas de IP₆K de mamífero en diferentes sistemas (bacteria y baculovirus).
- Puesta a punto de protocolos de purificación de las diferentes isoformas.
- Diseño de estrategias de cristalización de esta enzima: obtención de construcciones proteicas específicas y cristalización de complejos proteína-proteína.
- Obtención de información estructural de la IP₆K de mamífero que ayude a entender su función.

4.1.3 Materiales y métodos

4.1.3.1 Clonación y diseño de construcciones

Los genes *ip6k1*, *ip6k2* e *ip6k3* que codifican para las isoformas IP₆K1, IP₆K2 e IP₆K3 de humano se obtuvieron de clones comerciales. Para la expresión bacteriana de las mismas, el ADNc, en función del

caso, se insertó en los vectores pKLSLt (Angulo et al., 2011), pOPTG y pOPTM (cedidos por la Dra. Olga Perisic, MRC-LMB) o pGEX-4T (GE Healthcare) que permiten la expresión de la muestra de interés en fusión con las proteínas LSLt, GST y MBP respectivamente. Los ADNcs se amplificaron por PCR usando los oligonucleótidos 1-5 (Tabla 4.1), y se insertaron en el plásmido por clonación de restricciones. Para expresar IP₆K1 e IP₆K3 en células de insecto, *hip6k1* e *hip6k3* ADNc se clonaron en el vector de transferencia pAC8-G_DsRed mediante PCR usando los oligonucleótidos 6-7 indicados en la Tabla 4.1. A partir de estos vectores se produjeron los bácmidos necesarios para transfectar las células de insecto. La generación de los bácmidos se realizó por el grupo del Dr. Arnaud Potterszman (Instituto de Genética y Biología Molecular y Molecular de Estrasburgo, Francia), en el marco de un proyecto INSTRUCT que obtuvimos para la realización de este trabajo. El protocolo de generación de baculovirus utilizado es similar al reportado en el primer trabajo de investigación presentado en esta tesis (ver trabajo de investigación 1, sección 3.1.1).

Para el trabajo realizado también se produjeron una serie de construcciones de IP₆K2 correspondientes a versiones truncadas y mutadas, cuyos datos y oligonucleótidos (8-16) se recogen en la Tabla 4.1.

Tabla 4.1. Construcciones realizadas de IP₆K y respectivos oligonucleótidos utilizados.

Nº	Secuencia	Motivo	Construcción
1F	GGAATTCATATGTGTGTTGTCAAACCATGG	Introducir sitio NdeI	IP ₆ K1/pOPTG
1R	CGCGGATCCCTACTGGTTCCTCGTCCCGCATC	Introducir sitio BamHI	IP ₆ K1/pKOPTG
2F	GAGATATACATATGAGCCAGCCTTCAGG	Introducir sitio NdeI	IP ₆ K2/pOPTG y PoPTM
2R	CATCAGGATCCTCACTCCCACTCTCCA	Introducir sitio BamHI	IP ₆ K2/pOPTM
3F	CTATAGAATTCATGAGCCAGCCTTCAGG	Introducir sitio EcoRI	IP ₆ K2/pKLSL
3R	CATCAAAGCTTTCACCTCCCACTCTCCA	Introducir sitio HindIII	IP ₆ K2/pKLSL
4F	CCTGTATTTCAGGGCGAATTCATGGTTGTGCAAAACAGCGCAG ATG	"Megaprimer"	IP ₆ K3/pKLSL
4R	TGGTGCTCGAGTGGCGGCTCATTCTCCCTCTTGGATATC	"Megaprimer"	IP ₆ K3/pKLSL
5F	CGGATCTGGTTCGCGTGGATCCATGGTTGTGCAAAACAGCGCA GATG	"Megaprimer"	IP ₆ K3/pGEX-4T2
5R	GTCAGTCAGTCAGATGCGGCCCTCATCTCTCCCTCTTGGATA TC	"Megaprimer"	IP ₆ K3/pGEX-4T2
6F	GTTCCAGGGGCCCCATATGTGTGTTGTCAAACCATGG	Introducir sitio NdeI	IP ₆ K1/pAC8-G
6R	GCCCGGAATTCGGATCCCTACTGGTTCCTCGTCCCG	Introducir sitio BamHI	IP ₆ K1/pAC8-G
7F	GTTCCAGGGGCCCCATATGTGTGCAAAACAGCGC	Introducir sitio NdeI	IP ₆ K3/pAC8-G
7R	GCCCGGAATTCGGATCCCTCATCTCTCCCTCTTGGATATCC	Introducir sitio BamHI	IP ₆ K3/pAC8-G
8F	CTATAGAATTCGCCAAAGGCGTCTTCTGG	Introducir sitio EcoRI	Δ(1-20)-IP ₆ K2/PKLSL
8R	ATAGTTTAGCGGCGCTCACTCCCACTCTCCA	Introducir sitio NotI	Δ(1-20)-IP ₆ K2/PKLSL
9F	GAGATATACATATGGCCAAAGGCGTCTTCTGG	Introducir sitio NdeI	Δ(1-20)-IP ₆ K2/pOPTH
9R	CGCGGATCCGCCAAAGGCGTCTTCTGG	Introducir sitio BamHI	Δ(1-20)-IP ₆ K2/pOPTH
10F	GATAATTCAGACTGTGAACCAAAAAATTTATCTTACTGGAAAA CCTG	Mutagénesis dirigida	Δ(107-202)-IP ₆ K2/PKLSL
10R	CAGGTTTTTCAGTAAGATAAATTTTTTGGTTTACAGTCTGAAT TATC	Mutagénesis dirigida	Δ(107-202)-IP ₆ K2/PKLSL
11F	CTTTGAAGAAGATGAAGACAGGAACAACAGTACAAATTTATCT TACTG	Mutagénesis dirigida	Δ(74-198)-IP ₆ K2/PKLSL
11R	CAGTAAGATAAATTTTACTGGTTGTCTGTCTTCATCTTCTT CAAAG	Mutagénesis dirigida	Δ(74-198)-IP ₆ K2 /PKLSL
12F	CCCAGAGTGGTCTGGACATCGGCGCCAGCTCTGTAG	Mutagénesis dirigida	Δ(347-371)-IP ₆ K2 /PKLSL
12R	CTACAGAGTGGCGCGATGTCCAGGACCACTTCGGG	Mutagénesis dirigida	Δ(347-371)-IP ₆ K2 /PKLSL
13F	CTGCTGGTCAATTTATGATGGCGGCGCAGCTCTGTAGATG	Mutagénesis dirigida	Δ(337-377)-IP ₆ K2 /PKLSL
13R	CATCTACAGAGTGGCGCGCCATCATAAATGACCAGCAG	Mutagénesis dirigida	Δ(337-377)-IP ₆ K2/PKLSL
14F	CTTGTAACACATGTAGCGGCGCGGGAATATAAGTTCCAG	Mutagénesis dirigida	EKK164-166AAA-IP ₆ K2 /PKLSL
14R	CTGGGAACCTTATATCCCCGCCCGCTACAGTGTAGTACAAG	Mutagénesis dirigida	EKK164-166AAA-IP ₆ K2 /PKLSL
15F	TGAAGAGCCATAAGTTAGCAGCAGCATTTGAGTGGCTAAAG	Mutagénesis dirigida	EEE147-149AAA-IP ₆ K2/ PKLSL

15R	CTTTAGCCCACTCAATGCTGCTGCTAACTTATGGCTCTTCA	Mutagénesis dirigida	EEE147-149AAA-IP ₆ K2/ PKLSL
16F	CTGGGAACCTTATATTTCCCGCCGCTACAGTGTAGTACAAG	Mutagénesis dirigida	KK117/118AA-IP ₆ K2 /PKLSL
16R	CTGTTTCTAAGACATGATGTGCTGCGTTTGTGTCCACCTTAG	Mutagénesis dirigida	KK117/118AA-IP ₆ K2 /PKLSL

4.1.3.2 Expresión y purificación de IP₆K

4.1.3.2.1 Expresión y purificación de IP₆K recombinante en bacteria

La expresión de las tres isoformas de IP₆K con diferentes etiquetas, se llevó a cabo en células de *E. coli* Rosetta 2 (DE3) en medio LB. Cuando la densidad óptica (OD₆₀₀) de los cultivos a 37°C alcanzó un valor de 0,8-0.9 se indujo la expresión de la proteína recombinante con IPTG 1mM, y se mantuvo la expresión durante 20 horas a 16°C. Posteriormente, se recogieron los cultivos celulares por sucesivos ciclos de centrifugación a 4200 rpm durante 20 min, y se almacenaron a -80°C hasta su uso.

Se observó un mejor crecimiento en las muestras de IP₆K2 con etiqueta LSL, por lo que la purificación se ha optimizado para esta construcción. Los precipitados celulares se resuspendieron en el tampón de lisis A (20 mM Tris-HCl pH 8, 150 mM NaCl, 1mM DTT y DNAsa). Tras la centrifugación, la fracción soluble se cargó en una columna preparada manualmente de *Sefarosa* CL-6B preequilibrada en el tampón de lisis. Después de un lavado exhaustivo con el mismo tampón, la LSL-IP₆K2 se eluyó añadiendo 0.2M lactosa al mismo. Tras bajar su fuerza iónica mediante dilución, la muestra fue inyectada en una columna de heparina previamente equilibrada en el tampón B (20 mM Tris-HCl pH 8, 50 mM NaCl y 1mM DTT). Después de un lavado se realizó un gradiente entre el tampón B y C (20 mM Tris-HCl pH 8, 1M NaCl y 1mM DTT). La etiqueta LSL se eliminó mediante un corte con la proteasa TEV (relación 1:40, preparada en el laboratorio) agitando suavemente en cámara fría durante 16 horas. En este paso se añadió IP₆ 1mM, el cual se mantuvo a lo largo de toda la purificación. Con el fin de eliminar restos de LSL-IP₆K2 que permaneció sin cortar, la muestra se mezcló de nuevo con resina *Sefarosa* CL-6B equilibrada en buffer de lisis A con IP₆ 1mM. Por último, se realizó una cromatografía de exclusión molecular (Superdex 200 16/60), con el fin de eliminar posibles agregados, equilibrada previamente en el tampón de lisis A con IP₆ 1mM.

4.1.3.2.2 Expresión y purificación de IP₆K recombinante en células de insecto

La expresión de IP₆K1 e IP₆K3 en células de insecto se ha realizado en el grupo del Dr. Arnaud Poterzsmann. La expresión se evaluó en células SF21 y SF9, con tiempos de expresión de 48 y 72h. Finalmente se seleccionaron las células SF21 y 72 h de expresión. El protocolo de expresión utilizado es similar al reportado en el primer trabajo de investigación presentado en esta tesis (ver trabajo de investigación 1, sección 3.1.1). Los precipitados celulares se resuspendieron en el tampón de lisis D (20 mM Tris-HCl pH 8, 250 mM NaCl, 1mM IGEPAL, 2mM BME, 0,1mM PMSF, DNAsa). A continuación, las células se lisaron mediante ciclos de sonicación. El extracto celular se centrifugó durante 45 min a 17500 rpm a 4°C. La fracción soluble se filtró (0,45 µm) y se redujo su fuerza iónica mediante dilución 1:5 en buffer sin sal. A continuación, se introdujo la muestra en una columna de heparina de 5ml (GE Healthcare) previamente equilibrada en buffer B (20 mM Tris-HCl pH 8, 50 mM NaCl, 2mM DTT). Tras un lavado con el mismo tampón se eluyó la muestra con gradiente de sal entre el tampón B y el C (20 mM Tris-HCl pH 8, 1M NaCl, 2mM DTT). A continuación, el pico obtenido se inyectó en una columna de GST de 5ml debidamente equilibrada en el tampón E (20 mM Tris-HCl pH 8, 250 mM NaCl, 2mM DTT, 1mM IP₆). La proteína de fusión se eliminó mediante corte con la

proteasa PreScission (preparada en el laboratorio) en la columna de GST agitando suavemente durante 4 horas a 4°C. El último paso cromatográfico consistió en una columna de exclusión molecular Superdex 200 16/60 equilibrada en tampón E. Únicamente la muestra IP₆K1 presentó un rendimiento suficiente para continuar su estudio.

4.1.3.3 Obtención de nanoanticuerpos específicos de IP₆K2

4.1.3.3.1 Generación de nanoanticuerpos frente a IP₆K2

El grupo del Dr. Jan Steyaert, del Centro de Biología Estructural de Bruselas, se llevó a cabo la generación y selección de “*nanoanticuerpos*” (NBs, del inglés “*nanobodies*”) frente a la proteína completa IP₆K2 gracias a la financiación de un proyecto INSTRUCT que facilita el uso de plataformas de Biología Estructural en Europa.

Los NBs se generaron a través de la inmunización de la llama (*Lama glama*) con muestras puras de IP₆K2 humana completa en presencia de IP₆. El proceso completo dura aproximadamente unos seis meses. La inmunización de la llama se realizó a través de 5-6 inyecciones, una por semana, de 0,5 mg de proteína pura y después se extrajo el suero a partir del cual se obtuvieron los diferentes ADNc de los NBs. A continuación, éstos ADNc se introdujeron en un sistema de presentación de fagos (“*phage display*”) para proceder a la selección de los NBs. Después de la selección, los ADNcs se clonaron en el vector pMES4. A partir de este proceso se generaron 15 plásmidos, conteniendo NBs específicos para IP₆K2, en el vector pMES4. Este vector permite la expresión del NB incluyendo una señal pelB de exportación al espacio periplásmico en el extremo N-terminal, y una etiqueta de 6 histidinas en su extremo C-terminal.

Además de la proteína completa IP₆K2 con IP₆, se utilizaron otras construcciones en el proceso de selección de los NBs: IP₆K2 sin IP₆, una forma trunca de la proteína ($\Delta(107-202)$ -IP₆K2) que carece del inserto I) y la etiqueta de fusión LSL, con el fin de obtener mayor información durante el proceso de selección.

4.1.3.3.2 Expresión y purificación de los nanoanticuerpos

La expresión de los NBs se llevó a cabo en la cepa Wk6 que incluye resistencia a estreptomicina. Tras la transformación con los correspondientes plásmidos se prepararon preinóculos en medio LB-amp suplementado con 2% de glucosa y 1mM de MgCl₂. Se dejaron crecer toda la noche a 37°C en agitación. Una vez crecido, cada cultivo se transfirió al medio LB suplementado con glucosa y MgCl₂, 0.1% y 1mM respectivamente. Cuando el cultivo a 37°C alcanzó una densidad óptica OD₆₀₀= 0,7-1,2 se indujo la expresión con IPTG 1mM. Después de la inducción, las células se incubaron en agitación a 25°C durante toda la noche. Finalmente, los cultivos celulares se centrifugaron a 4200 rpm durante 20 min para preparar los precipitados celulares que se conservaron a -80°C hasta ser procesados.

La purificación de los NBs es relativamente sencilla puesto que se expresan en el espacio periplásmico. La lisis se realizó mediante resuspensión de las células en medio TES durante una hora en frío y a continuación durante 45 min también en frío, en medio TES/4. Después se centrifugó durante 30 min a 8000 rpm y se guardó el sobrenadante (fracción periplásmica) e introdujo en una columna de heparina previamente equilibrada en tampón fosfato F (50 mM Sodio fosfato pH 7 y 1M NaCl). A continuación,

se lavó con el tampón fosfato F y después con el tampón fosfato G (50 mM Sodio fosfato pH 6 y 1M NaCl). Por último, se realizó un gradiente entre los tampones G y H (50 mM Sodio fosfato pH 7, 150 mM NaCl y 0.3 M Imidazol pH 7). De forma general, el rendimiento de la purificación varía de 1 a 20 mg/l dependiendo de cada NB.

4.1.3.4 Ensayos de cristalización de IP₆K, recogida y análisis de datos

La búsqueda inicial y la optimización de las condiciones de cristalización se llevó a cabo tal y como se detalla en el anexo de Materiales y Métodos (ver sección 8.3.5). Se obtuvieron cristales en los experimentos realizados con las muestras de IP₆K1 expresada en células de insecto en los ensayos de cristalización preliminares.

La recogida de datos de difracción se realizó en el sincrotrón de ALBA en la línea de macromoléculas XALOC. Los datos se indexaron, integraron y escalaron con los programas XDS y Aimless (ver sección 8.5 del anexo de Materiales y Métodos). Se utilizaron métodos convencionales de resolución de estructura mediante reemplazamiento molecular (MOLREP, PHASER), utilizando como modelo inicial la estructura de otras IPKs resueltas (1W2C, 2IEW, 2XAN, 4O4E).

4.1.3.5 Identificación de las especies cristalizadas

Los cristales de IP₆K1 obtenidos se analizaron mediante técnicas electroforéticas (SDS-PAGE). Inicialmente, se lavó un cristal de proteína de IP₆K1 en una solución estabilizante y posteriormente se disolvió en 2µl de agua para cargarlo en un gel poliacrilamida al 12%. El contenido de la gota de cristalización también fue aplicado en el gel. Después se realizó la tinción con plata del gel siguiendo el protocolo establecido.

Por otro lado, se preparó un gel para secuenciación del N-terminal. Para ello se disolvió nuevamente un cristal de proteína en agua, que posteriormente se cargó en un gel de poliacrilamida al 15%. A continuación, el gel se transfirió a una membrana (*Immobilon-P Millipore*) que se tiñó rápida y brevemente para asegurar la transferencia de la banda de interés. Por último, se recortó la banda cuidadosamente y se envió a un servicio de secuenciación del N-terminal de proteínas.

4.1.3.6 Modelado de IP₆K2

Para el análisis de secuencias de las distintas isoformas de IP₆K humana y la homología entre ellas y con otras IPKs se ha llevado a cabo una combinación de varios programas: *seaview* y *pymol* se han utilizado para realizar los alineamientos estructurales de las proteínas, *ESPrpt* para la representación de estos alineamientos, *JPred* y *Psipred* para la predicción de estructura secundaria y *pymol* para representar y analizar los modelos estructurales obtenidos.

El modelado por homología de secuencia (“*homology modeling*”) es la técnica más utilizada en el modelado de proteínas y presupone que proteínas que presentan un alto porcentaje de identidad en secuencia también comparten una estructura similar. Los servidores SWISS-MODEL (<https://swissmodel.expasy.org>) y Phyre-2 (“*Protein Homology/analogY Recognition Engine*”) (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>), buscan, a partir de la secuencia de

interés, proteínas homologas resueltas en las bases de datos (PDB) para utilizar como modelos en la predicción de su estructura tridimensional. Estos servidores funcionan por alineamientos de secuencia con proteínas homólogas con altos porcentajes de identidad de secuencia, por lo que, si la secuencia de la estructura a modelar presenta una identidad menor al 20% con las estructuras disponibles, el modelado resulta menos eficiente. Además, el servidor Phyre-2 incorpora una simulación de plegamiento “*ab initio*” para modelar regiones de proteínas sin homología detectable con las estructuras conocidas. La familia de IP₆Ks poseen largas inserciones que son específicas de su familia (insertos I y II). En particular, el inserto I posee más de un centenar de residuos que han sido eliminados (residuos 78-198) para la elaboración del modelo, ya que interfieren en la búsqueda de proteínas homólogas. Así, para el modelado, se ha utilizado la secuencia de la IP₆K2 humana que carece del inserto I. Por otro lado, se ha analizado el inserto I de manera independiente, con el fin de establecer si hay homólogos fuera de la familia de las IPKs para esta región.

4.1.4 Resultados y discusión

4.1.4.1 Producción de IP₆K2 en *E. coli* y cristalización

Se ha llevado a cabo la expresión de las tres isoformas de IP₆Ks en células bacterianas con etiquetas de fusión (LSL, GST y MBP). Sin embargo, la isoforma IP₆K1, se produce en su mayoría agregada, y la IP₆K3 es altamente insoluble. Por el contrario, la isoforma IP₆K2 presenta una buena expresión en la cepa Rosetta2, produciendo los mejores resultados con la etiqueta LSL.

El primer paso de purificación de IP₆K2 expresada en células bacterianas mediante la columna de *Sefarosa* CL-6B permite aislar con éxito la proteína LSL-IP₆K2 (67.8 kDa) (Figura 4.1a, carril 1). El segundo paso de purificación con columna de heparina incrementa el grado de pureza de LSL-IP₆K2 (Figura 4.1a, carril 2). La proteína eluye en torno a un 55-60% del tampón A, correspondiente a 550-600 mM de NaCl. Después del corte con TEV, que es prácticamente total, se vuelve a pasar por una pequeña cantidad de resina *Sefarosa* CL-6B con el fin de eliminar posible IP₆K2 no cortada (Figura 4.1a, carriles 3, 4 y 5). El último paso de purificación consiste en una cromatografía de exclusión molecular en el que se obtiene un único pico de proteína que eluye en el ml 71, compatible con una especie monomérica de IP₆K2 (Figura 4.1b).

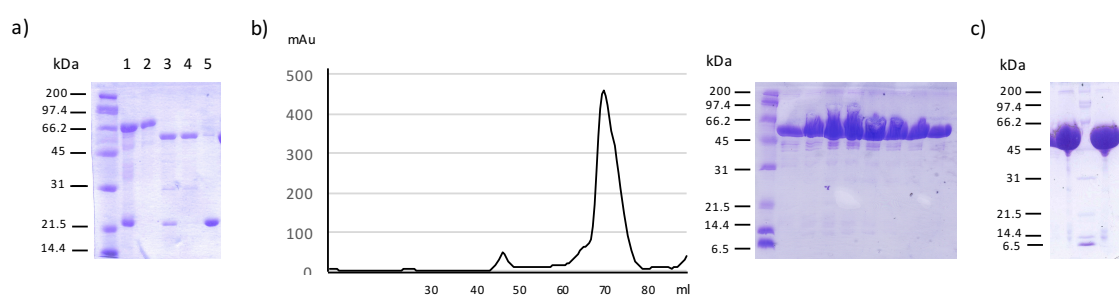


Figura 4.1. Purificación de IP₆K2. a) SDS-PAGE al 12%, carril 1: LSL-IP₆K2 obtenida tras el paso por *Sefarosa* CL-6B (67.8 kDa), carril 2: LSL-IP₆K2 obtenida tras la columna de heparina, carril 3: IP₆K2 después del corte con TEV (49.18 kDa), carril 4: IP₆K2 después de pasar por segunda vez por *Sefarosa*, carril 5: resina *Sefarosa* con LSL retenida y LSL-IP₆K2 no cortada. b) Perfil de la cromatografía de exclusión molecular de IP₆K2 y fracciones del pico obtenido. c) SDS-PAGE al 12% de la muestra final con la que se llevó a cabo la cristalización de la enzima. Carril de la izquierda, muestra sin IP₆; carril de la derecha, muestra con IP₆.

Por cada litro de cultivo se obtienen aproximadamente 7.5 mg totales de proteína pura. La muestra final se concentra hasta 12-18 mg/ml. La purificación sin IP₆ se realizó siguiendo el mismo protocolo detallado anteriormente (Figura 4.1c), aunque el rendimiento obtenido disminuye a la mitad y no es posible concentrar la muestra más de 3 mg/ml.

Se realizaron múltiples ensayos de cristalización con la proteína completa de IP₆K2. En ningún caso esta muestra produjo cristales de proteína. Por ello, se diseñaron una serie de construcciones correspondientes a formas truncadas y/o mutadas de la proteína IP₆K2 (Figura 4.2). En resumen, se han producido construcciones que carecen de los primeros residuos del N-terminal Δ 1-20, del inserto I del N-terminal: Δ 107-202 y Δ 74-198, y del inserto II del C-terminal: Δ 348-370 y Δ 338-372 (Tabla 4.1). Además, se han producido construcciones que carecen de grupos (“clusters”) de residuos básicos agrupados en el N-terminal, con el fin de reducir la entropía de la superficie: cluster 1: KK117-118AA, cluster 2: EEE147-149AAA, y cluster 3: EKK164-166AAA, además de una combinación de los clusters 1 y 3, y del Δ 348-370/cluster 3. La purificación de estas construcciones se realizó siguiendo el protocolo descrito anteriormente. En general, los rendimientos son más bajos que el obtenido con la proteína nativa. A excepción del mutante Δ 338-372 que muestra mucha tendencia a la precipitación, todos se comportan de manera similar a la nativa.

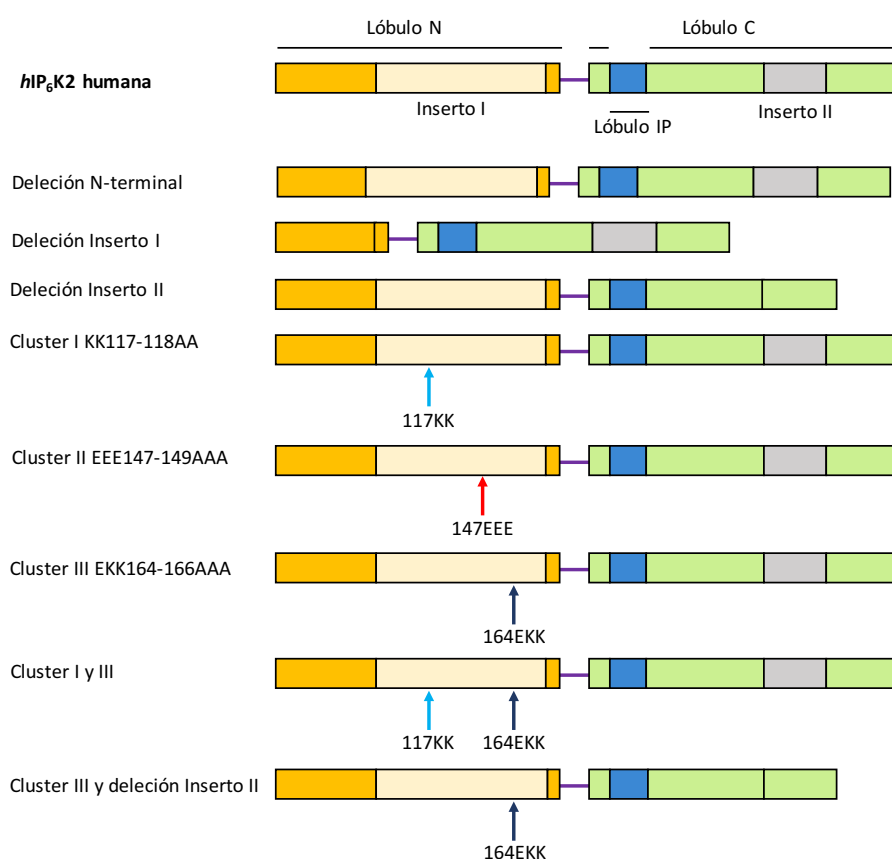


Figura 4.2. Diferentes construcciones realizadas sobre hIP₆K. En naranja se representa el lóbulo N-terminal con el inserto I en amarillo, en morado el conector, en verde el lóbulo C-terminal con el inserto II en gris y en azul el lóbulo de unión a IP.

Desafortunadamente, ninguna de las construcciones expresadas en *E. coli* ha permitido obtener cristales de esta proteína. Por tanto, se decidió llevar a cabo dos aproximaciones experimentales para favorecer la cristalización de alguna de las isoformas de IP₆K: 1) expresión de las isoformas IP₆K1 e IP₆K3 en células de insecto utilizando el sistema de baculovirus y 2) cocristalización de la isoforma IP₆K2 con NBs específicos.

4.1.4.2 Producción de IP₆K1 en células de insecto y cristalización

Inicialmente, se llevaron a cabo pilotos de expresión de las dos isoformas en células de insecto: IP₆K1 e IP₆K3. La isoforma IP₆K1 se expresó notablemente mejor que la IP₆K3 por lo que se continuó trabajando con esta isoforma.

El primer paso de purificación de IP₆K1 consistió en una columna de heparina que permitió “limpiar” sustancialmente la proteína de fusión GST-IP₆K1 (77 kDa) (Figura 4.3a). Posteriormente, el paso por una columna de afinidad GST seguido por el corte con la proteasa PreScission en la propia columna de GST incrementa significativamente su grado de pureza (Figura 4.3b). A la vista del perfil obtenido en un último paso de cromatografía de exclusión molecular (Figura 4.3c) se concluye que hay mezcla de especies, posiblemente correspondiente con las formas monomérica/dimérica de IP₆K1.

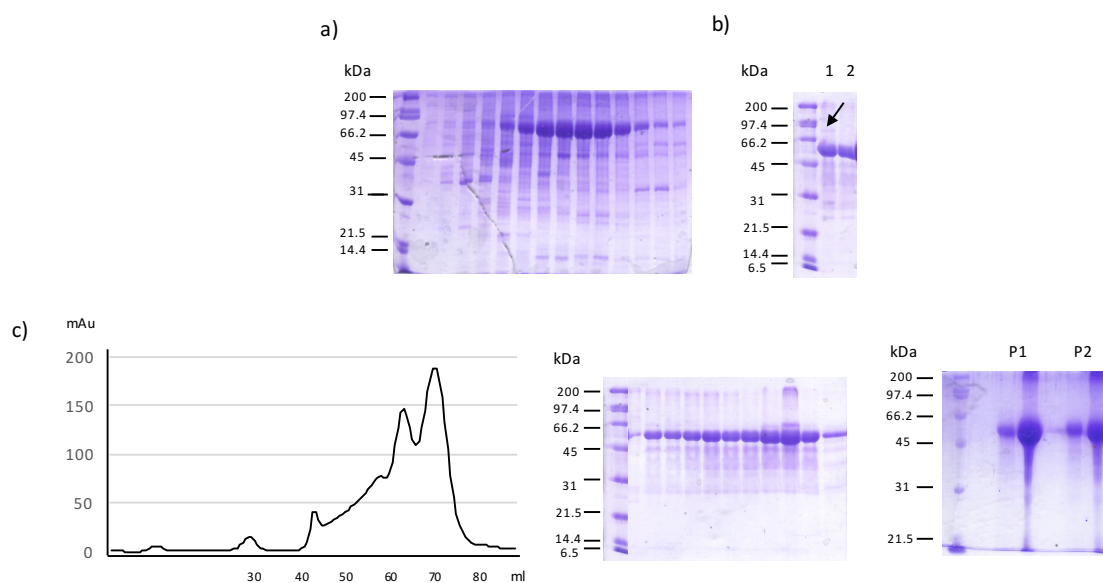


Figura 4.3. Purificación de IP₆K1. a) SDS-PAGE del 12% mostrando la GST-IP₆K1 (77 kDa) obtenida tras el paso de captura con la columna de heparina. Se observan las fracciones correspondientes a la elución con el gradiente de sal entre 50 mM y 1 M. b) SDS-PAGE del 12%. Los carriles muestran, 1: proteína correspondiente a la elución de la columna de GST después del corte con PreScission (50.5 kDa). Se aprecia una pequeña banda (señalada con flecha) correspondiente a GST-IP₆K1 sin cortar, 2: Después de pasar la muestra por una segunda columna de GST la banda de proteína sin cortar desaparece. c) Pasos finales de purificación de IP₆K1. A la izquierda, perfil de la cromatografía de exclusión molecular de IP₆K1. En el centro, SDS-PAGE del 12% con las fracciones del pico obtenido. A la derecha, SDS-PAGE al 12% de la muestra final de IP₆K1 durante el proceso de concentración con la que se llevó a cabo la cristalización de la enzima, P1 y P2 indican el primer y segundo picos obtenidos en la exclusión molecular correspondiente al dímero y monómero respectivamente.

A partir de las células SF21 con un tiempo de expresión de 72h se obtienen 4,5 mg por litro de cultivo. La especie monomérica representa un 62% del total, y la especie dimérica alrededor del 34%. Cuando el primer paso cromatográfico se realiza con una columna de GST en lugar de la columna de heparina, el rendimiento total de la purificación disminuye significativamente.

Los diferentes picos se concentraron por separado y se utilizaron en ensayos independientes de cristalización. La búsqueda de condiciones de cristalización iniciales se realizó a diferentes concentraciones, 3,5 mg/ml, 5,5 mg/ml y 12.5 mg/ml, mediante técnicas de alto rendimiento (ver anexo Materiales y Métodos). Si bien el rendimiento de la purificación es bajo, se obtuvo material para la realización de ensayos iniciales de cristalización utilizando casi 400 condiciones comerciales. Se obtuvieron cristales muy pequeños (dimensiones aproximadas de $5\mu\text{m} \times 4\mu\text{m}$) en diversas condiciones de cristalización y que tardaron más de dos meses en aparecer. La mayoría de estos cristales presentan forma de placa, y una gran anisotropía en el patrón de difracción. El mejor cristal obtenido se obtuvo en la condición de cristalización: 0.2M Fosfato de sodio/potasio, 0.1M Bis-Tris-Propano pH 7.5, 20% PEG 3350 y difractó a una resolución máxima de 1.6 Å (Figura 4.4). A partir de este cristal, conseguimos obtener un conjunto de datos completo a 1.8 Å (Figura 4.4c). Las estadísticas del procesado se muestran en la Tabla 4.2.

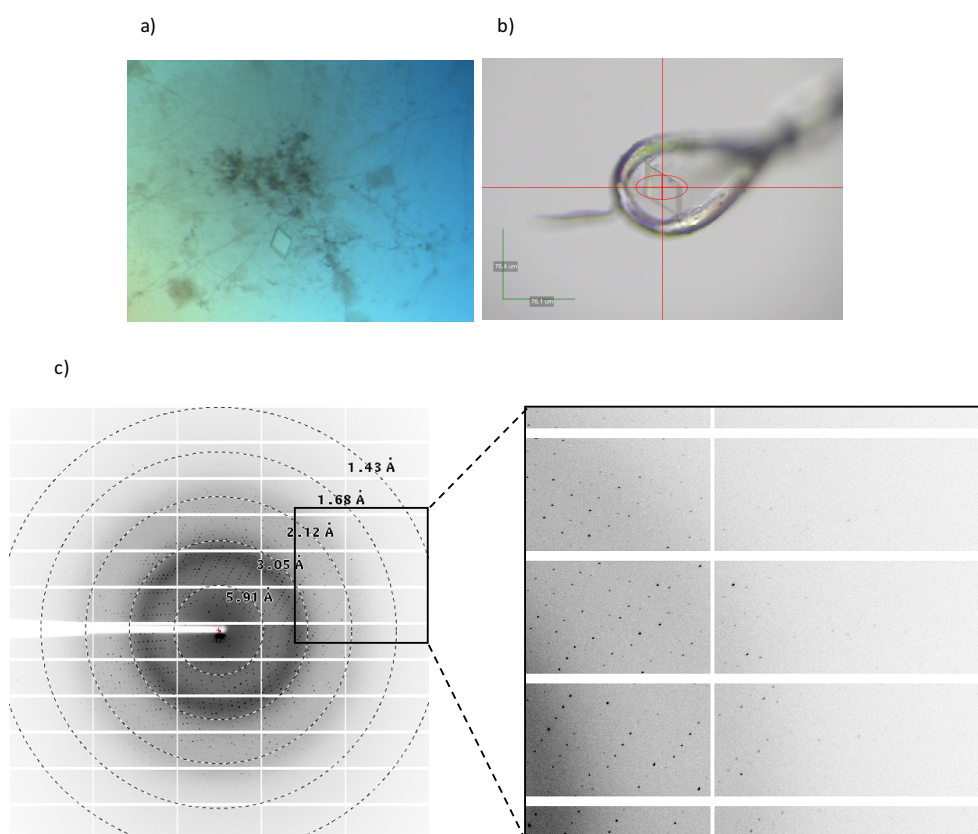


Figura 4.4. Cristales y patrones de difracción de IP₆K1. a) Imagen del mejor cristal obtenido de IP₆K1 en la gota de cristalización, b) Cristal montando en un lazo de nylon en el Sincrotrón ALBA y c) patrón de difracción del cristal.

	IP ₆ K1 Baculovrius
Grupo espacial	I 121
Celdilla unidad	
<i>a</i> (Å)	88.3
<i>b</i> (Å)	94.1
<i>c</i> (Å)	112.0
α (Å)	90.0
β (Å)	95.9
γ (Å)	90.0
T (K)	100
Fuente de radiación	Sincrotrón
Longitud de onda (Å)	0.979490
Intervalo de resolución (Å)	47.04-1.80 (1.83-1.80)
Reflexiones totales	564220 (30358)
Reflexiones únicas	84163(4464)
<i>I</i> / σ (<i>I</i>)	14.7 (2.6)
Compleitud	99.8 (100.0)
Multiplicidad	6.7 (6.8)
R _{pim}	3.5 (3.4)
R _{merge}	7.6 (83.3)
CC 1/2	99.8 (71.9)

Tabla 4.2. Estadísticas del procesado. Los valores entre paréntesis hacen referencia a la última capa de resolución.

A pesar de los múltiples intentos llevados a cabo, hasta la fecha no se ha conseguido resolver la estructura de la especie cristalizada por reemplazo molecular. Se han utilizado como modelos la estructura completa resuelta de IP₆K de *Entamoeba histolytica* (EhIP₆K) (código PDB 4O4C) y sus dominios por separado, así como diferentes estructuras de las IPKs conocidas (*HsIP*₃ 3-K: 1W2C, *AtIP*₅ 2-K: 2XAN, *ScIPmK*: 2IEW). Además, se han intentado formar derivados de átomo pesado sin éxito con sales de sodio (NaI), oro (KAuCn₂) y platino (Pt).

El hecho del largo periodo necesario para la aparición de los cristales (dos meses) sugiere que la proteína podría haberse degradado en las condiciones de cristalización. Para comprobarlo, se analizó la gota de cristalización en la que apareció el mejor cristal obtenido en un gel de acrilamida al 12%. El gel demuestra que no hay banda correspondiente a la proteína completa (50.5 kDa) y en su lugar se observan dos bandas de tamaño aproximado a 40 kDa y unos 18 kDa (Figura 4.5a). Para averiguar cuál de las dos bandas corresponde con los cristales obtenidos se analizó un cristal de IP₆K2 en un gel de acrilamida al 12%, y posteriormente se realizó una tinción con plata del gel (Figura 4.5b). Como se ve en el gel, la banda corresponde con un tamaño de unos 18 kDa por lo que de esta manera se confirma que los cristales obtenidos de IP₆K1 corresponden con la banda de menor tamaño observada en la gota de cristalización. Con el fin de conocer a qué fragmento de la IP₆K1 podría corresponder la especie cristalizada, se hicieron varios intentos de secuenciación del N-terminal, pero debido a la poca cantidad de muestra no se obtuvieron resultados concluyentes (Figura 4.5c).

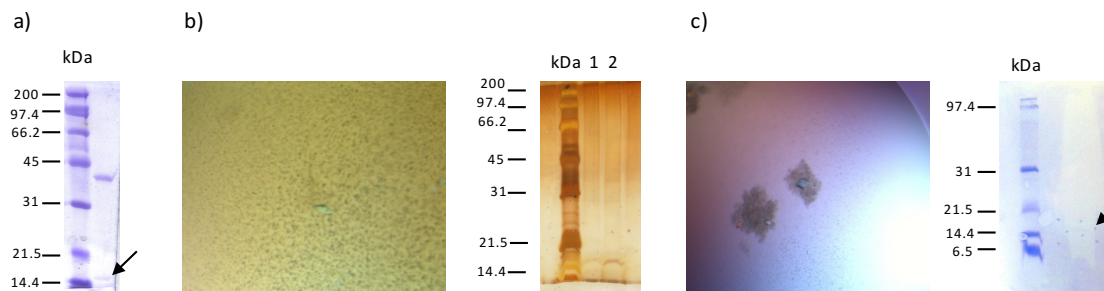


Figura 4.5. Análisis de los cristales de IP₆K1. a) Análisis mediante SDS-PAGE 12% de la gota de cristalización, se observa una banda de unos 40 kDa y otra inferior indicada con una flecha de unos 18 kDa. b) A la izquierda, cristal de IP₆K1 utilizado para teñir con plata y a la derecha Gel SDS-PAGE 12% con tinción de plata. Carril 1: Cristal disuelto en agua, carril 2: Resto de la gota de cristalización. Ambos carriles muestran una banda de unos 18 kDa. c) A la izquierda cristales utilizados para transferir a membrana, a la derecha, imagen de la membrana teñida a la que se transfirió un gel del 15% donde previamente se había cargado un cristal de IP₆K1. Se observa una pequeña banda por encima de los 14 kDa que se recortó y mandó a secuenciar.

Por el momento, se continúa trabajando en la resolución estructural de la especie cristalizada. El proceso es muy complicado debido a la baja reproducibilidad de estos cristales junto al desconocimiento de la secuencia del fragmento cristalizado y a la limitación de la muestra de la que disponemos.

4.1.4.3 Producción de nanoanticuerpos y formación de complejos con IP₆K2

Inicialmente se seleccionaron un total de 15 NBs mediante el procedimiento detallado anteriormente (Figura 4.6).

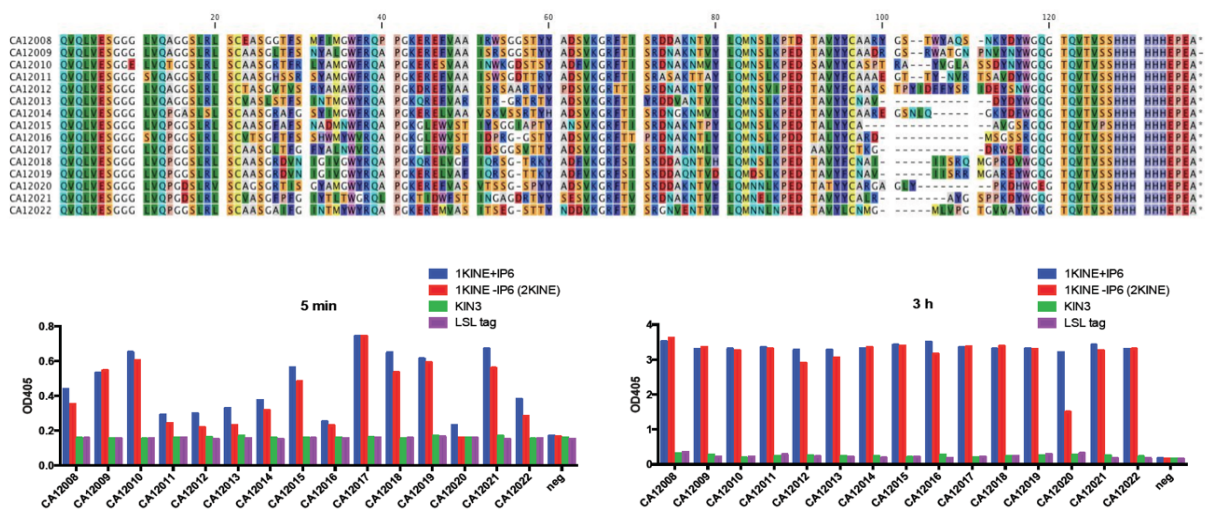


Figura 4.6. Selección de NBs. a) Secuencia de los NBs seleccionados para llevar a cabo la formación de complejos con IP₆K2. b) Resultados de los experimentos de ELISA después de 5 min (izquierda) y 3 horas (derecha). En azul IP₆K2 con IP₆, en rojo IP₆K2 sin IP₆, en verde la forma truncada de la proteína (Δ 95-IP₆K2) y en morado la etiqueta de fusión LSL.

La purificación de los diferentes NBs consiste en un único paso utilizando una columna de níquel, que permite aislar las muestras con elevada pureza (Figura 4.7).

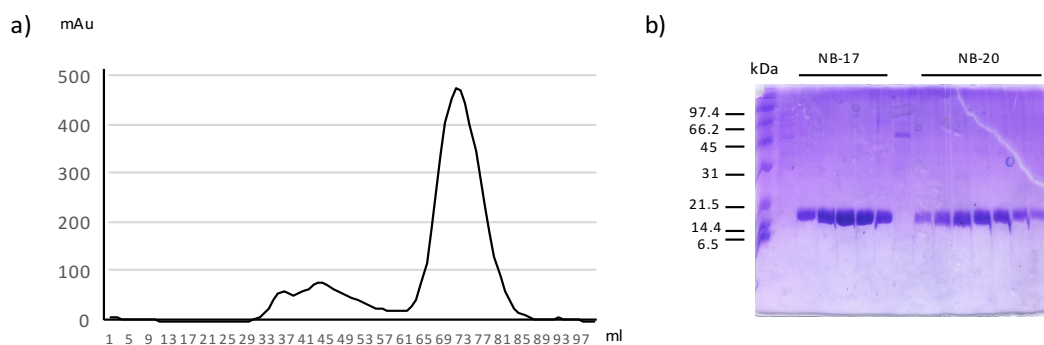


Figura 4.7. Purificación de los NB. a) Perfil de elución de la purificación del NB 20 mediante cromatografía de afinidad a níquel. b) Análisis mediante SDS-PAGE del 15% de los picos obtenidos en la cromatografía de afinidad de los NB-17 (13.56 kDa) y -20 (14.35 kDa).

Para la formación del complejo entre cada NB y la IP₆K2, se mezclaron ambas muestras, previamente concentradas al volumen deseado, en un exceso molar de NB 3:1. La mezcla se incubó en hielo unos 30 min y después se realizó una cromatografía de exclusión molecular utilizando una columna Superdex 75 16/600. Posteriormente mediante SDS-PAGE se evaluó la formación del complejo correspondiente. (Figura 4.8).

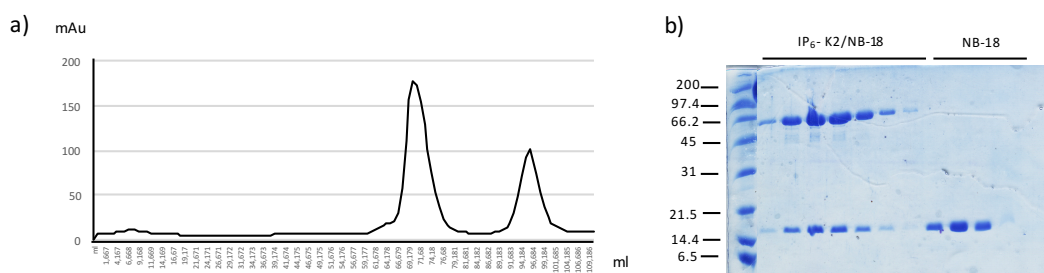


Figura 4.8. Formación del complejo IP₆K2/NB. a) Estudio de la formación del complejo de la IP₆K2 con el NB-18, el primer pico corresponde con el complejo IP₆K2/NB-18 (Volumen de elución= 70,3 ml), el segundo con el exceso de NB-18 (Volumen de elución= 95,9 ml). b) Análisis electroforético del complejo SDS-PAGE al 15% de acrilamida. El primer pico, que corresponde al complejo IP₆K2/NB-18, y el segundo con el exceso del NB-18.

De todos los NBs aislados, únicamente NB-18 fue capaz de formar el complejo “*in vitro*” con IP₆K2 (Figura 4.8). Con el complejo IP₆K2/NB-18 formado se ha realizado un muestreo masivo de condiciones de cristalización. Sin embargo, hasta la fecha, no se han producido cristales.

El hecho de que, únicamente, uno de los quince NBs investigados haya formado complejo con IP₆K2 sugiere que la selección de NBs llevada a cabo no ha sido adecuada. En la actualidad se está realizando una nueva selección en el grupo del Dr. Steyaert y a corto plazo se espera obtener anticuerpos específicos que permitan continuar con esta estrategia. Mediante este trabajo se ha puesto a punto el protocolo de expresión y purificación de estos NBs y la formación de sus respectivos complejos con IP₆K2.

Figura 4.9. Alineamiento de enzimas de la familia de las IPKs. Se incluyen las siguientes IPKs: *EhIP₆K*, *HsIP₆K1*, *HsIP₆K2*, *HsIP₆K3*, *ScIPmK*, *HsIPmK* y *HsIP₃ 3-K* (*Eh*: *Entamoeba histolytica*; *Hs*: *Homo sapiens*; *Sc*: *Saccharomyces cerevisiae*). En sombreado naranja y morado se muestran las inserciones del N-terminal de *EhIP₆K* y *HsIPmK* respectivamente. En sombreado verde el inserto I y en amarillo el inserto II de *HsIP₆K1*.

4.1.4.4.2 Modelado de IP₆K2 y análisis estructural

Se ha obtenido un modelo de IP₆K2 mediante el uso de dos servidores de modelado por homología distintos. Por un lado, el servidor Phyre-2 ha seleccionado la estructura de IPmK de *S. cerevisiae* (código PDB 2IF8) para realizar el modelo y, por otro lado, el servidor SWISS-MODEL ha utilizado la estructura de *EhIP₆K* (código PDB 4O4C). En ambos casos, se utilizó como entrada la secuencia de IP₆K2 sin el inserto I puesto que este dificulta el alineamiento con las bases de datos estructurales y por tanto la generación de un modelo fiable. Los dos modelos generados son prácticamente iguales, si bien muestran una mayor diferencia en relación con el inserto II. A continuación, se muestra el modelo generado por Phyre-2 (Figura 4.10a). Una superposición del modelo obtenido con la estructura de la única IP₆K conocida hasta la fecha perteneciente a *E. histolytica* (*EhIP₆K*) y con la estructura de IPmK humana (*hIPmK*) se muestra en la Figura 4.10b. Se ha utilizado la estructura de IPmK de humano puesto que es la única estructura de esta familia que se ha conseguido obtener en presencia de inositol.

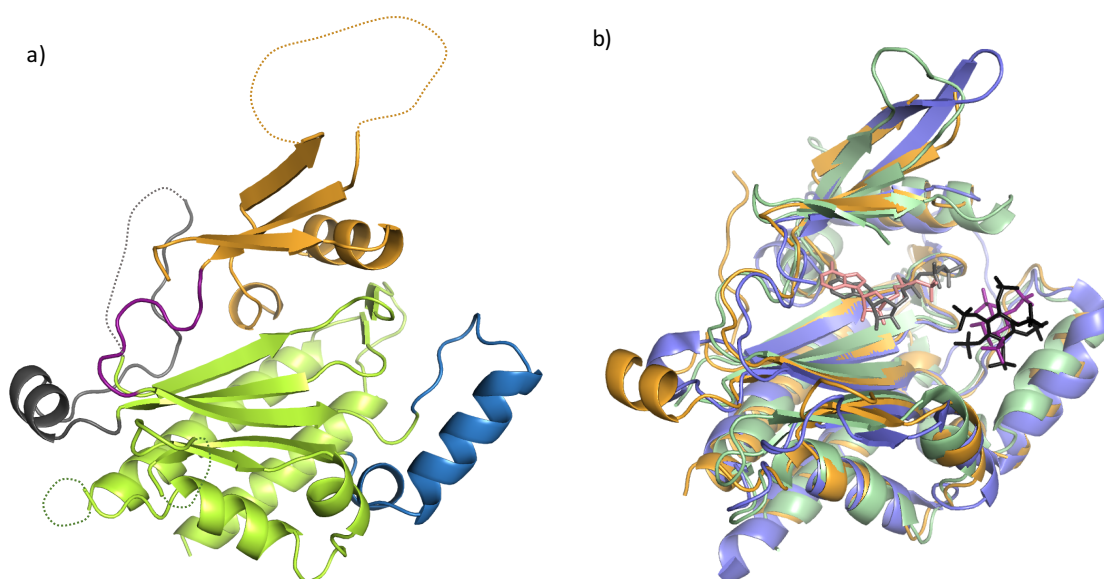


Figura 4.10. a) Modelo de IP₆K2 generado por Phyre-2. En diagrama de cintas se representa el lóbulo N-terminal en naranja, el lóbulo C-terminal en verde y el lóbulo IP en azul. El conector entre los lóbulos N y C-terminal aparece en violeta y el inserto II en gris. En línea de puntos se muestran las zonas no trazadas en el modelo. **b) Comparación del modelo de IP₆K2 con *EhIP₆K* y *hIPmK*.** En naranja se representa la estructura del modelo generado de IP₆K2, en azul la estructura de *EhIP₆K* (código PDB 4O4C) con el IP₆ en negro y el ATP en gris, y en verde la estructura de *hIPmK* (código PDB 5W2H) con el IP₃ en magenta y el ADP en salmón.

A partir del alineamiento de secuencia y del modelo obtenido, se puede postular qué residuos estarían involucrados en la coordinación del sustrato y el reconocimiento del ATP. En relación con la coordinación del inositol, podemos observar como las IP₆Ks y las IPmKs únicamente presentan una hélice en el lóbulo IP (Figura 4.11). El modelo sugiere que los siguientes residuos de esta hélice podrían intervenir en la coordinación del IP: Lys115, Gln119 y Lys122 (Figura 4.11). Residuos equivalentes coordinan el sustrato en *EhIP₆K2* y *hIPmK* (Figura 4.11). Otras regiones de lóbulo C también aportan

residuos en la unión del inositido (Lys101, Lys153 y Arg157), bastante conservados en la familia. Por otro lado, en las dos enzimas usadas como modelo se observa participación del lóbulo N en la coordinación del inositido: *EhIP₆K* muestra una inserción en el N-terminal, que coordina el IP₆ través de la Gly23 y Gly27 (Figura 4.11), mientras que *hIPmK* muestra una corta inserción en el N-terminal que también coordina el inositido a través de la Arg82 (Figura 4.11). El alineamiento estructural muestra la localización de estas dos inserciones en *EhIP₆K* e *hIPmK* y se observa que no están presentes en *hIP₆K* (Figuras 4.9 y 4.12). Sin embargo, la isoforma humana muestra una larga inserción (inserto I) que probablemente esté involucrada en el reconocimiento del inositido. De hecho, la enzima más distante de la familia de las IPKs, IP₅ 2-K, posee una inserción en esta misma región bastante más reducida que es esencial en la coordinación del N-terminal.

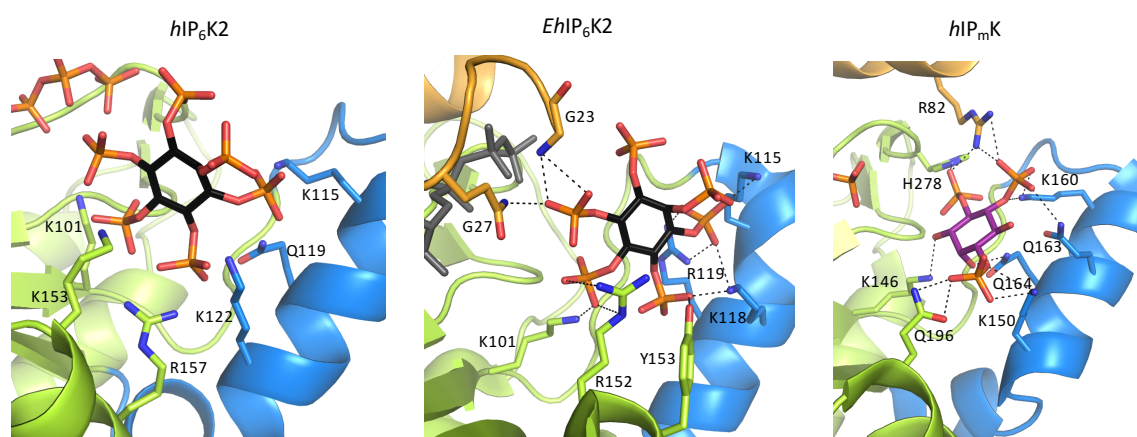


Figura 4.11. Coordinación del sustrato en *hIP₆K2*, *EhIP₆K* y *hIPmK*. A la izquierda, el modelo obtenido para la *hIP₆K2*, en el centro la estructura de *EhIP₆K* y a la derecha *hIPmK*. Se mantiene el código de colores establecido: el lóbulo N en naranja, el lóbulo C-terminal en verde y el IP en azul. El IP₆ se muestra en negro y el IP₃ en magenta, destacándose los átomos fosfato en naranja y los oxígenos en rojo.

Respecto al reconocimiento del nucleótido, los residuos están absolutamente conservados en posición con todas las IPKs y las PKs. Los residuos clave en esta interacción mediante puentes de hidrógeno serían la Lys42, Glu86, Leu88, Asp99, Asp262, y Lys101 (Figura 4.12). También hay un reconocimiento hidrofóbico del anillo de adenina y ribosa que se encuentra conservado.

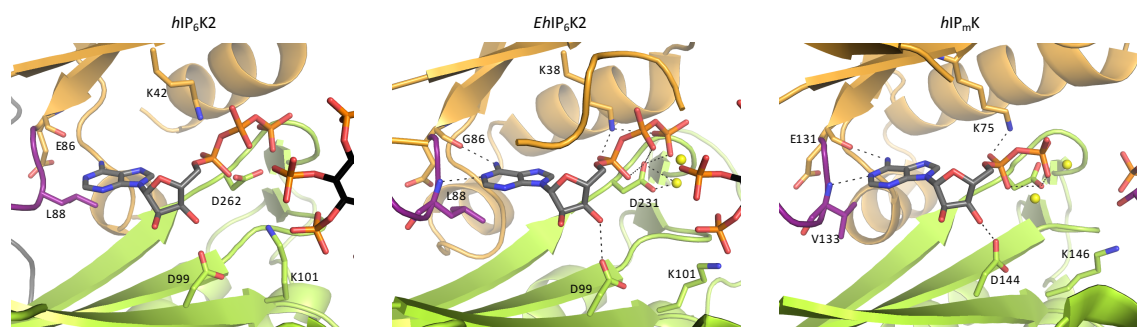


Figura 4.12. Reconocimiento del nucleótido en *hIP₆K2*, *EhIP₆K* y *hIPmK*. A la izquierda, el modelo obtenido para la *hIP₆K2*, en el centro la estructura de *EhIP₆K* y a la derecha *hIPmK*. Se mantiene el código de colores establecido anteriormente, el nucleótido se representa en gris y los átomos de magnesio en esferas amarillas. Se representan los residuos polares involucrados en el reconocimiento del nucleótido. Los residuos hidrofóbicos que forman parte del bolsillo de unión a la adenina y ribosa no se representan, pero están conservados en posición en todos los casos.

Por otro lado, y para obtener información sobre la región específica de la familia de las IP₆Ks, se ha intentado modelar el inserto I utilizando los servidores mencionados (SWISS-MODEL y Phyre-2), pero debido a que no existe ninguna proteína con suficiente identidad de secuencia los modelos generados no son fiables. Por tanto, se predice que este fragmento de la proteína, insertado en el N-terminal, podría presentar un plegamiento nuevo. Como se ha visto anteriormente, la implicación del N-terminal en el reconocimiento del sustrato es frecuente en las IPKs.

En relación con el inserto II de 45 residuos, el modelo de IP₆K2 inicialmente mostrado predice una hélice pequeña y un bucle muy largo (representado en gris en la Figura 4.10a). No obstante, dado que en función del programa utilizado la estructura propuesta para este inserto II varía, la fiabilidad del modelo generado es muy baja en esta zona. Merece la pena destacar que la IPmK de *S. cerevisiae*, posee también un inserto localizado en la misma región que el II de IP₆K2, con un elevado contenido de residuos de ácido aspártico (Figura 4.9). Esta inserción de IPmK produce interacciones proteína-proteína, lo que nos hace pensar que el inserto II de las diferentes IP₆Ks también podría estar vinculada a esta función.

4.1.5 Conclusiones IP₆Ks

Las principales conclusiones obtenidas de este trabajo se detallan a continuación:

- La isoforma IP₆K2 es soluble cuando se expresa en células bacterianas y se obtiene un buen rendimiento en su purificación. La isoforma IP₆K1 se produce utilizando células de insecto en condiciones adecuadas, aunque el rendimiento no es muy elevado.
- Las muestras de IP₆K2 presentan desorden que impide su cristalización, probablemente centrado en las grandes inserciones que posee con respecto a las otras IPKs.
- Las muestras de IP₆K1 se digieren en las condiciones de cristalización generando cristales de un fragmento de 18 kDa que difracta a una resolución máxima de 1.8 Å.
- El alineamiento de IP₆K de humano con otras IPKs de estructura conocida muestra conservación únicamente en pequeños motivos importantes para reconocimiento de sustratos y catálisis.
- El modelo estructural de IP₆K2 ha permitido predecir los residuos posiblemente involucrados en la unión del nucleótido y del inositol. No se ha podido modelar la región del inserto I debido a que no conserva secuencia con otras proteínas.
- La técnica de NBs podría ayudar a la cristalización de IP₆K2 al unir regiones flexibles y por tanto fijar conformaciones determinadas.

4.1.6 Perspectivas

Este trabajo ha supuesto un avance significativo en el estudio estructural de IP₆K a la vez que inicia diferentes vías para continuar en el futuro:

Por un lado, hay que insistir en la identificación de la forma cristaliza de IP₆K1. El hecho de que ningún fragmento estructural conocido de la familia de las IPKs haya funcionado en el reemplazamiento molecular, sugiere que este fragmento podría tratarse del inserto I cuyo plegamiento es desconocido. Para abordar esta opción, nos planteamos la generación de construcciones del inserto del N-terminal, lo que puede hacerse bien mediante experimentos de proteólisis limitada o bien mediante biología

molecular. Por otro lado, y considerando que el principal objetivo es cristalizar la proteína IP₆K completa, nos proponemos explorar a fondo la vía ya iniciada de NBs que es una alternativa muy prometedora. En los últimos años se han reportado numerosos casos en los que este método ha sido clave para la obtención de cristales de proteínas especialmente problemáticas.

Finalmente, se han identificado residuos relevantes para la coordinación del sustrato y la actividad de la proteína. La realización de experimentos de mutagénesis dirigida apoyaría estos resultados. Además, sería muy interesante también la realización de experimentos de mutagénesis dirigida en residuos básicos del inserto I y valorar así su impacto en la unión a IP. Esto permitiría la identificación de nuevos residuos implicados en la actividad de la enzima.

5. DISCUSIÓN GENERAL

5. Discusión general

Como se ha puesto de manifiesto a lo largo de este trabajo de tesis, cada vez es mayor el número de trabajos científicos que describen las interferencias entre la ruta de los IPs y el metabolismo del ARN. Mediante el empleo de la Cristalografía de rayos X y diferentes técnicas bioquímicas se ha logrado profundizar en el conocimiento de proteínas que intervienen en ambas rutas: por un lado, se han estudiado las IPKs de mamífero que intervienen en la regulación de los niveles de los IPs y, por otro, proteínas de unión a ARN y en concreto Nrd1. Entender el funcionamiento de este tipo de enzimas es clave para la comprensión de los sofisticados mecanismos que regulan los procesos celulares.

La figura 5.1 trata de recoger en un esquema las interacciones descritas hasta la fecha entre ambas rutas (IPs y ARN). Estas interacciones se encuentran en muchos puntos de la ruta de los IPs, desde el control de su precursor PIP₂, hasta la producción de los IPPs, los últimos IPs sintetizados en la ruta. Muchas de las proteínas representadas en la figura 5.1 poseen papeles muy estudiados en cáncer, apoptosis, viabilidad celular, regulación de la transcripción, biogénesis del ARN, etc. El esquema mostrado probablemente es limitado, puesto que hoy en día falta conocimiento para poder completarlo.

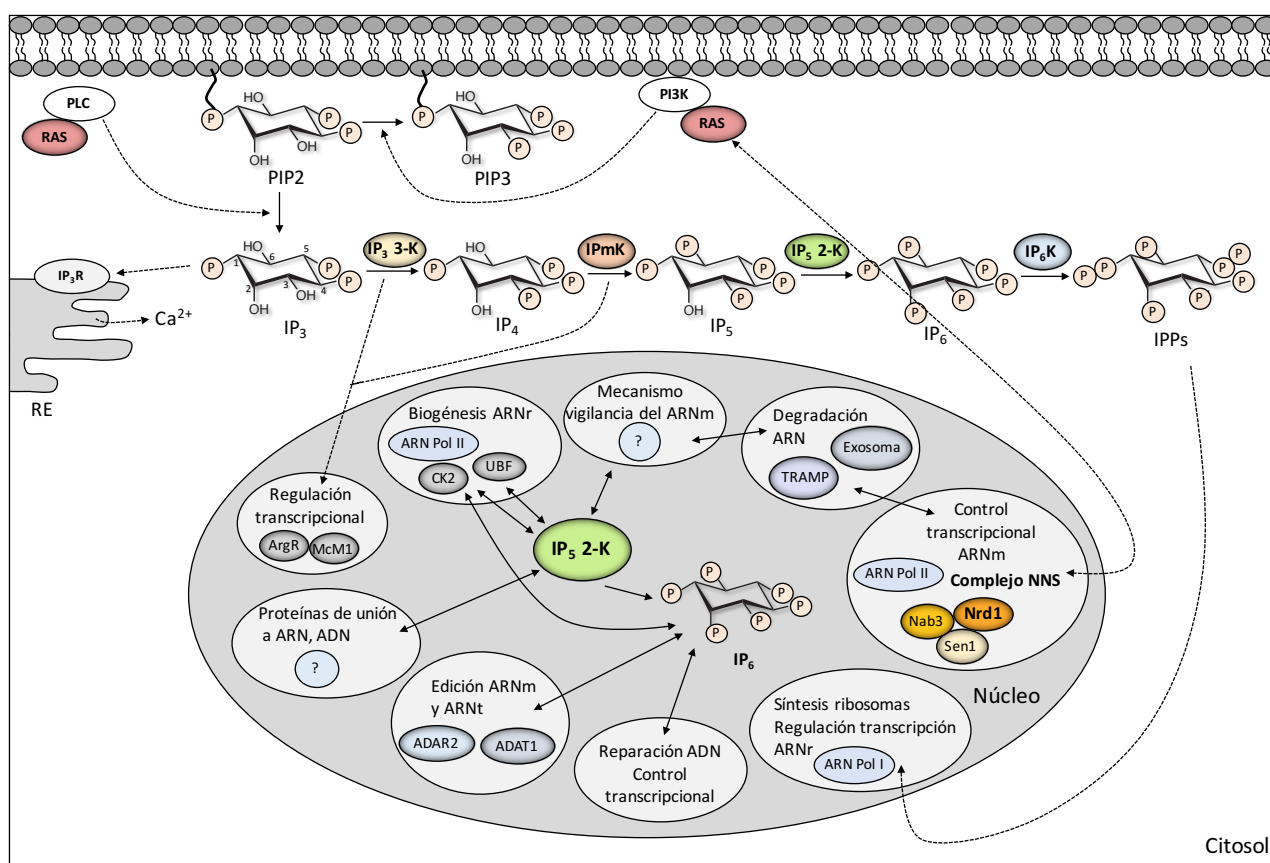


Figura 5.1. Esquema global de las rutas de los IPs/IPKs y relación con el metabolismo del ARN. Se detalla de forma esquemática el inicio de la ruta de los IPs y las diferentes reacciones de fosforilación por las IPKs, así como las principales funciones de los IPs/IPKs en el núcleo, citosol celular, retículo endoplasmático (RE) y su relación con las vías del ARN.

El comienzo de la ruta de los IPs aúna a proteínas y segundos mensajeros esenciales en señalización celular (RAS, PI3K, PIP₃ e IP₃). Estas proteínas y compuestos son necesarios para la supervivencia celular, cada uno jugando un papel específico. En particular, RAS y PI3K son dos dianas hoy en día muy perseguidas por la industria farmacéutica en la lucha contra el cáncer. Del mismo modo, el resto de las proteínas y moléculas representadas en la figura 5.1, poseen funciones esenciales. En particular, dos enzimas de la familia de las IPKs implicadas en la regulación de los IPs que han sido estudiadas en esta tesis (IP₅ 2-K e IP₆K) muestran también una estrecha relación con patologías severas como el cáncer. Así, la IP₆K ayuda al control del cáncer induciendo apoptosis, y como se comentará más adelante, la expresión de IP₃ 3-K aumenta el potencial metastásico de las células cancerígenas. Por otro lado, se han reportado papeles anticancerígenos para algunos de los IPs, y en particular para el IP₆, el cual como se ha detallado en esta tesis es sintetizado por la IP₅ 2-K. El amplio espectro de funciones de estas IPKs se debe a que poseen efectos pleiotrópicos: son proteínas multifuncionales con actividad catalítica, mediante su actividad quinasa, y reguladora, mediante la interacción con otras proteínas de señalización celular y biogénesis del ARN.

La Biología Estructural, campo en el que se enmarcan los estudios realizados en esta tesis, plantea amplios objetivos que abarcan desde el conocimiento del funcionamiento básico de las macromoléculas biológicas hasta el diseño de fármacos o inhibidores de su función basándose en datos estructurales y moleculares. En este último aspecto, un caso paradigmático es el diseño racional de inhibidores de las proteínas quinasas (PKs) basados en su estructura. Las PKs son protagonistas destacadas en señalización celular ya que fosforilan y activan a otras proteínas, actuando así como interruptores moleculares de rutas de señalización. Por tanto, son dianas moleculares muy perseguidas para la terapia de múltiples enfermedades. La estrategia más desarrollada se basa en el boqueo de la unión del ATP inhibiendo así la función quinasa de la PK. En algunos casos, se han conseguido inhibidores potentes que ya han sido comercializados como *Gleevec*, que inhibe la tirosina quinasa Bcr-abl y se emplea principalmente en el tratamiento de la leucemia mieloide crónica. Es posible por tanto desarrollar inhibidores muy selectivos para una quinasa. No obstante, se han identificado más de 520 PKs (que constituyen lo que se denomina como el “quinoma” humano) (Manning et al., 2002), todas las cuales reconocen el ATP de modo similar. Esto conlleva efectos colaterales no deseados, ya que un inhibidor puede bloquear enzimas diferentes a la diana de interés. Las IPKs, como se ha mencionado, constituyen un subgrupo de las PKs. Si bien estas enzimas fosforilan derivados de inositol en lugar de a otras proteínas, el mecanismo catalítico es similar y el reconocimiento de ATP se conserva. El desarrollo de inhibidores específicos contra las enzimas IPKs es también un campo de máximo interés por sus potenciales aplicaciones clínicas y por su uso en la caracterización funcional de estas enzimas. Ya se han reportado algunos inhibidores específicos de las IPKs basados en el bloqueo de la unión de ATP y, por ejemplo, se sabe que el inhibidor TNP (N2-(m-(trifluorometil) benzil) N6-(p- nitrobenzil) purina) un análogo de purina que imita al ATP, es capaz de inhibir tanto la IP₃ 3-K como la IP₆K1 (Padmanabhan et al., 2009). Debido a su naturaleza, es muy posible que este inhibidor lo sea también de alguna de las PKs del quinoma humano. No obstante, y gracias a los trabajos aquí presentados junto con los de otros grupos de investigación, hoy en día disponemos de información muy específica del reconocimiento del sustrato IP en casi todas las IPKs. Por tanto, el diseño de inhibidores dirigidos al sitio de unión a IP minimiza los problemas de interferencia con otras enzimas. En la actualidad, se ha identificado un inhibidor competitivo del sitio de unión a inositol de la IP₃ 3-K, BIP-4 (Schroder et al., 2015), el cual disminuye el potencial metastásico de células de cáncer de pulmón por lo que es un compuesto prometedor para uso terapéutico. Estudios estructurales llevados a cabo en nuestro grupo de investigación han revelado los posibles modos de unión de este ligando (Schroder et al., 2015), y se

planea seguir estos estudios para ayudar a mejorar las características de este inhibidor. Respecto a la IP₅ 2-K, la inhibición de la isoforma según la especie de la que procede puede tener intereses muy distintos. Mientras la inhibición de la isoforma de mamífero es de interés en el área de la salud, la isoforma de planta presenta un elevado interés agronómico, ya que se persigue el desarrollo de cereales con bajo contenido en IP₆ para evitar problemas de malnutrición en países en vías de desarrollo y problemas de contaminación de las aguas. Curiosamente, la inhibición que se persigue en la isoforma de hongo posee otros intereses bien distintos, y también en el área de la salud. Así, el hongo patógeno *Cryptococcus neoformans* es causante de meningoencefalitis en humanos y la IP₅ 2-K contribuye a su virulencia, ya que esta se ve reducida significativamente por la delección de CnIP₅ 2-K (Oh et al., 2017). Por tanto, la inhibición de esta enzima podría ayudar a combatir la infección.

Pero no solamente la inhibición de estas proteínas ha de centrarse en el bloqueo del sitio activo del dominio catalítico, ya que poseen en su mayoría otros dominios reguladores y además interaccionan con otras proteínas implicadas en funciones muy diversas. Las interacciones proteína-proteína son importantes en muchos procesos biológicos, ya que intervienen en el control del ciclo celular, metabolismo, diferenciación, plegamiento de proteínas, señalización celular, transcripción, modificaciones postraduccionales y transporte. Existen varios ejemplos en la bibliografía que demuestran la importancia de la inhibición de rutas bloqueando este tipo de interacciones (Mansilla et al., 2017). En el caso de las IPKs, merece la pena mencionar la regulación de IP₃ 3-K mediante una interacción directa con CaM. La información sobre datos estructurales y bioquímicos obtenidos en nuestro grupo de investigación han permitido avanzar en la comprensión de las bases moleculares que definen la interacción de estas dos proteínas (Franco-Echevarria et al., 2014). Es necesario explorar las posibles consecuencias a nivel celular del bloqueo de esta interacción. Por otro lado, el dominio de unión a F-actina (ABD; “*Actin-binding Domain*”) del extremo N-terminal de IP₃ 3-K, al igual que el dominio catalítico, es también esencial para promover la metástasis en células cancerígenas (Windhorst et al., 2011). Por tanto, la ruptura de la interacción ABD/F-Actina podría ser otra estrategia de inhibición de la función de IP₃ 3-K. En este aspecto, aún queda pendiente descifrar la interacción entre el dominio ABD de IP₃ 3-K con la F-actina y por ello recientemente se ha iniciado esta línea de investigación en el grupo de investigación donde se ha desarrollado esta tesis. Otro ejemplo lo constituye la IP₅ 2-K de mamífero, cuya primera estructura se ha determinado en esta tesis. Como se ha comentado, la IP₅ 2-K interacciona con enzimas implicadas en la biogénesis del ARNr (CK2, TCOF1 y UBF) y, además, interfiere con uno de los mecanismos de vigilancia del ARNm. La estructura ha revelado la presencia de un parche básico en su superficie muy marcado, conservado en las isoformas de mamífero y que podría estar relacionado con estas funciones. De hecho, la presencia de parches básicos en la superficie de las proteínas se relaciona con la unión a ADN/ARN, membranas y fosfolípidos, o con otras proteínas por complementariedad de carga. Existen ejemplos en la literatura de cada caso en particular. Así, la proteína Dmc1 humana (homólogo de RAD51), una recombinasa de ADN específica de la meiosis, forma un anillo octamérico que presenta parches básicos internos y externos en su estructura a los que se une el ADN (Kinebuchi et al., 2004). El fosfolípido PIP₃ regula el tráfico de membranas interactuando con el dominio FYVE, el cual contiene superficies básicas de unión específica al IP₃ (Misra y Hurley, 1999). Finalmente, la interacción entre dominios de reclutamiento de dos caspasas (CARD o “*caspase recruitment domains*”), RAIDD e ICH-1 está mediada a través del parche básico de RAIDD y el parche ácido de ICH-1 (Chou et al., 1998). El próximo objetivo de este trabajo incluye el estudio en detalle de las interacciones de IP₅ 2-K con las proteínas que interacciona. Esto proporcionaría información muy útil sobre las posibles funciones adicionales de la enzima y, a su vez, cuales son los puntos clave para el diseño de moléculas inhibitoras de esta interacción. Finalmente, la

IP₆K, cuyo estudio estructural también se ha abordado en esta tesis, también interacciona con otras proteínas y en particular con la chaperona HSP90. En este caso, esta interacción sí estaría relacionada con la función catalítica de la enzima, ya que se sabe que bloquea la síntesis de los IPPs. Se ha demostrado que la interacción entre ambas proteínas disminuye los niveles de IPPs implicados en apoptosis y, en consecuencia, promueve el desarrollo del cáncer (Koldobskiy y Snyder, 2011; Thota y Bhandari, 2015). En este caso el desarrollo de moléculas inhibidoras de esta interacción podría tener consecuencias favorables en el tratamiento de esta enfermedad. En definitiva, el conocimiento de todas las interfaces aportaría más datos para el diseño de inhibidores de las interacciones proteína-proteína o en su caso, proteína-ARN.

Considerando otro tipo de diana, diversos grupos de investigación están abordando en la actualidad la inhibición de la proteína RAS mediante el diseño de moléculas capaces de bloquear la conformación inactiva de la enzima, con el fin de paralizar la señalización que conduce a la división celular en células cancerígenas. De hecho, una de las líneas ya iniciadas en el grupo de investigación se centra en la formación de complejos RAS-ligando y su resolución estructural mediante Cristalografía de rayos X, con el fin de aportar datos relevantes para el diseño de inhibidores más potentes y específicos. En particular, se ha encontrado una relación muy interesante entre el complejo NNS (Nrd1-Nab3-Sen1) y la señalización mediante RAS, que parecen tener efectos antagónicos. En respuesta a condiciones de carencia de nutrientes, concretamente a una deficiencia de glucosa, Nrd1 y Nab3 facilitan la represión de genes requeridos para un rápido crecimiento como *TYE7*, *DLD3* y *CLN3*, genes involucrados en la glicólisis, el metabolismo, y la progresión del ciclo celular respectivamente. Se sabe que la inhibición de la terminación de la transcripción es el resultado de la fosforilación de algún componente del complejo NNS por efectores de la vía RAS. Se han identificado dos sitios de fosforilación diferentes en Nab3, uno de ellos corresponde a un sitio de consenso de la PKA (una proteína quinasa y uno de los principales efectores de RAS). Por tanto, se postula que la proteína Nab3 pueda ser una diana directa de la PKA. A pesar de que se tiene un mayor conocimiento de este complejo mecanismo, todavía quedan muchos interrogantes por resolver (Darby et al., 2012).

En conclusión, el análisis estructural de las IPKs presentado en esta tesis puede contribuir de forma muy beneficiosa al diseño de inhibidores de esta familia de enzimas. No obstante, sería necesario tener una visión más detallada de la estructura de IP₆K, para lo cual se han propuesto diversas estrategias en esta tesis. Por otro lado, el conocimiento de la especificidad de unión a ARN de Nrd1 presentado ha permitido avanzar en el entendimiento del proceso de terminación de transcripción mediada por el complejo NNS, si bien otros aspectos como su interacción con la señalización mediada por RAS requieren de más trabajo futuro. Pese a que muchas relaciones entre la ruta de los IPs y el ARN están identificadas, entender cómo se producen estas interacciones mediante la biología estructural es fundamental para completar la visión que tenemos de estos sistemas.

6. CONCLUSIONES

6. Conclusiones

Se enumeran a continuación las principales conclusiones que se extraen del estudio de cada una de las proteínas caracterizadas en este trabajo.

Inositol pentaquisfosfato 2-quinasas (IP₅ 2-Ks):

1. La estructura de la proteína *mIP₅ 2-K* en presencia de inositido (*mIP₅ 2-K+IP₆* y *mIP₅ 2-K+IP₅+ATP*) revela que *mIP₅ 2-K* conserva el plegamiento general de las IPKs. Además, la forma cristalizada es equivalente a la conformación cerrada de *AtIP₅ 2-K*.
2. El reconocimiento de sustratos/productos está bastante conservado con el descrito para *AtIP₅ 2-K*, si bien presenta algunas diferencias significativas. Estas se centran en el reconocimiento de los fosfatos P1 y P3 del inositido y en la presencia de un nuevo residuo ácido catalítico que completa la coordinación del Mg²⁺ (Asp439).
3. Ambas isoformas de IP₅ 2-K (*mIP₅ 2-K* y *AtIP₅ 2-K*) presentan inserciones muy diferentes, dentro de las cuales cabe destacar la presencia de un novedoso sitio de zinc en *mIP₅ 2-K* fundamental para mantener tanto la estabilidad como su plegamiento.
4. *mIP₅ 2-K* muestra una elevada concentración de residuos positivamente cargados en su superficie configurando un gran parche básico que está conservado en las isoformas de mamífero. Se propone que esta región podría ser una interfaz implicada en las interacciones proteína-proteína o proteína-ARN relacionadas con la participación de *hIP₅ 2-K* en la biogénesis del ARNr.
5. Las estructuras obtenidas suponen una herramienta muy valiosa para ayudar en el diseño de inhibidores enzimáticos basados en el sitio de reconocimiento del inositido.

Inositol hexaquisfosfato quinazas (IP₆Ks):

6. La isoforma IP₆K2 se expresa muy satisfactoriamente en células bacterianas produciendo un buen rendimiento en su purificación y la isoforma IP₆K1 se expresa en muy pequeña cantidad pero sin agregar en células de insecto.
7. Múltiples experimentos de cristalización no han rendido cristales adecuados sugiriendo que las muestras de IP₆K2 tienen un desorden inherente que impide su cristalización, probablemente localizado en las largas inserciones que posee.
8. Hemos cristalizado un fragmento de 18 kDa de la IP₆K1 cuya estructura no es posible resolver utilizando modelos basados en otras IPKs. Por tanto, se sugiere que este fragmento podría corresponder con el inserto I y se está explorando esta posibilidad.
9. El modelo estructural de IP₆K2 obtenido a partir de modelado por homología es coherente con los alineamientos estructurales con otras IPKs y ha permitido identificar posibles residuos involucrados en la unión del nucleótido y del inositido.
10. Se propone que la técnica de cristalización con NBs constituye una alternativa razonable para abordar la cristalización de IP₆K2, mediante la fijación de ciertas zonas flexibles.

Nrd1:

11. La región de unión a ARN en Nrd1 está formada por dos dominios: RRM y SD. El dominio SD está definido por las regiones que flanquean la secuencia del RRM, generando una configuración nueva para los dominios RRM. La interfaz de interacción de los dos dominios, RRM y SD, es extensa y de carácter polar, y se mantiene mediante una red de tres argininas.
12. Las estructuras de esta región de Nrd1 con ARN (GUAA, CGUAAA y UUAGUAAUCC) revelan que ambos dominios, RRM y SD, están involucrados en el reconocimiento específico de unión a ARN, y únicamente se reconoce la secuencia GUAA de forma específica.
13. La estructura libre y en complejo con ARN han permitido establecer que el sitio de unión a ARN se encuentra preformado, con lo que la unión del ARN apenas produce cambios conformacionales. (Allen y Majerus, 1972)

7. BIBLIOGRAFÍA

7. Bibliografía

- Agarwal, R., Mumtaz, H., and Ali, N. (2009). Role of inositol polyphosphates in programmed cell death. *Mol Cell Biochem* 328, 155-165.
- Allen, R.H., and Majerus, P.W. (1972). Isolation of vitamin B12-binding proteins using affinity chromatography. 3. Purification and properties of human plasma transcobalamin II. *J Biol Chem* 247, 7709-7717.
- Angulo, I., Acebron, I., de las Rivas, B., Munoz, R., Rodriguez-Crespo, I., Menendez, M., Garcia, P., Tateno, H., Goldstein, I.J., Perez-Agote, B., *et al.* (2011). High-resolution structural insights on the sugar-recognition and fusion tag properties of a versatile beta-trefoil lectin domain from the mushroom *Laetiporus sulphureus*. *Glycobiology* 21, 1349-1361.
- Assenberg, R., Wan, P.T., Geisse, S., and Mayr, L.M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Curr Opin Struct Biol* 23, 393-402.
- Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., and Possee, R.D. (1994). The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. *Virology* 202, 586-605.
- Bacikova, V., Pasulka, J., Kubicek, K., and Stefl, R. (2014). Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res* 42, 8024-8038.
- Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* 10, 411-421.
- Bang, S., Kim, S., Dailey, M.J., Chen, Y., Moran, T.H., Snyder, S.H., and Kim, S.F. (2012). AMP-activated protein kinase is physiologically regulated by inositol polyphosphate multikinase. *Proc Natl Acad Sci U S A* 109, 616-620.
- Banos-Sanz, J.I., Sanz-Aparicio, J., Whitfield, H., Hamilton, C., Brearley, C.A., and Gonzalez, B. (2012). Conformational changes in inositol 1,3,4,5,6-pentakisphosphate 2-kinase upon substrate binding: role of N-terminal lobe and enantiomeric substrate preference. *J Biol Chem* 287, 29237-29249.
- Bennett, M., Onnebo, S.M., Azevedo, C., and Saiardi, A. (2006). Inositol pyrophosphates: metabolism and signaling. *Cell Mol Life Sci* 63, 552-564.
- Bergfors, T. (2003). Seeds to crystals. *J Struct Biol* 142, 66-76.
- Berridge, M.J., and Irvine, R.F. (1984). Inositol trisphosphate, a novel second messenger in cellular signal transduction. *Nature* 312, 315-321.
- Bhandari, R., Juluri, K.R., Resnick, A.C., and Snyder, S.H. (2008). Gene deletion of inositol hexakisphosphate kinase 1 reveals inositol pyrophosphate regulation of insulin secretion, growth, and spermiogenesis. *Proc Natl Acad Sci U S A* 105, 2349-2353.
- Boggon, T.J., and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure* 8, R143-149.
- Bozsik, A., Kokeny, S., and Olah, E. (2007a). Molecular mechanisms for the antitumor activity of inositol hexakisphosphate (IP6). *Cancer Genomics Proteomics* 4, 43-51.
- Bozsik, A., Kokeny, S., and Olah, E. (2007b). Molecular Mechanisms for the Antitumor Activity of Inositol Hexakisphosphate (IP subset6). *Cancer Genomics Proteomics* 4, 43-52.
- Brehm, M.A., Schenk, T.M., Zhou, X., Fanick, W., Lin, H., Windhorst, S., Nalaskowski, M.M., Kobras, M., Shears, S.B., and Mayr, G.W. (2007). Intracellular localization of human inositol 1,3,4,5,6-pentakisphosphate 2-kinase. *Biochem J*.
- Brondyk, W.H. (2009). Selecting an appropriate method for expressing a recombinant protein. *Methods Enzymol* 463, 131-147.

- Byrum, J., Jordan, S., Safrany, S.T., and Rodgers, W. (2004). Visualization of inositol phosphate-dependent mobility of Ku: depletion of the DNA-PK cofactor InsP6 inhibits Ku mobility. *Nucleic Acids Res* 32, 2776-2784.
- Carrasco, M.A., Marambio, P., and Jaimovich, E. (1997). Changes in IP3 metabolism during skeletal muscle development in vivo and in vitro. *Comp Biochem Physiol B Biochem Mol Biol* 116, 173-181.
- Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol* 24, 6241-6252.
- Chakraborty, A., Koldobskiy, M.A., Bello, N.T., Maxwell, M., Potter, J.J., Juluri, K.R., Maag, D., Kim, S., Huang, A.S., Dailey, M.J., *et al.* (2010). Inositol pyrophosphates inhibit Akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell* 143, 897-910.
- Chakraborty, A., Koldobskiy, M.A., Sixt, K.M., Juluri, K.R., Mustafa, A.K., Snowman, A.M., van Rossum, D.B., Patterson, R.L., and Snyder, S.H. (2008). HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. *Proc Natl Acad Sci U S A* 105, 1134-1139.
- Chakraborty, A., Werner, J.K., Jr., Koldobskiy, M.A., Mustafa, A.K., Juluri, K.R., Pietropaoli, J., Snowman, A.M., and Snyder, S.H. (2011). Casein kinase-2 mediates cell survival through phosphorylation and degradation of inositol hexakisphosphate kinase-2. *Proc Natl Acad Sci U S A* 108, 2205-2209.
- Chamberlain, P.P., Qian, X., Stiles, A.R., Cho, J., Jones, D.H., Lesley, S.A., Grabau, E.A., Shears, S.B., and Spraggon, G. (2007). Integration of inositol phosphate signaling pathways via human ITPK1. *J Biol Chem*.
- Chamberlain, P.P., Sandberg, M.L., Sauer, K., Cooke, M.P., Lesley, S.A., and Spraggon, G. (2005). Structural insights into enzyme regulation for inositol 1,4,5-trisphosphate 3-kinase B. *Biochemistry* 44, 14486-14493.
- Cheek, S., Ginalski, K., Zhang, H., and Grishin, N.V. (2005). A comprehensive update of the sequence and structure classification of kinases. *BMC Struct Biol* 5, 6.
- Cheung, J.C., Salerno, B., and Hanakahi, L.A. (2008). Evidence for an inositol hexakisphosphate-dependent role for Ku in mammalian nonhomologous end joining that is independent of its role in the DNA-dependent protein kinase. *Nucleic Acids Res* 36, 5713-5726.
- Chou, J.J., Matsuo, H., Duan, H., and Wagner, G. (1998). Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. *Cell* 94, 171-180.
- Clery, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* 18, 290-298.
- Communi, D., Dewaste, V., and Erneux, C. (1999). Calcium-calmodulin-dependent protein kinase II and protein kinase C-mediated phosphorylation and activation of D-myo-inositol 1,4, 5-trisphosphate 3-kinase B in astrocytes. *J Biol Chem* 274, 14734-14742.
- Communi, D., Vanweyenberg, V., and Erneux, C. (1997). D-myo-inositol 1,4,5-trisphosphate 3-kinase A is activated by receptor activation through a calcium:calmodulin-dependent protein kinase II phosphorylation mechanism. *EMBO J* 16, 1943-1952.
- Cox, M.M. (2012). Recombinant protein vaccines produced in insect cells. *Vaccine* 30, 1759-1766.
- Crocco, P., Saiardi, A., Wilson, M.S., Maletta, R., Bruni, A.C., Passarino, G., and Rose, G. (2016). Contribution of polymorphic variation of inositol hexakisphosphate kinase 3 (IP6K3) gene promoter to the susceptibility to late onset Alzheimer's disease. *Biochim Biophys Acta* 1862, 1766-1773.

Dailey, M.J., and Kim, S. (2012). Inositol polyphosphate multikinase: an emerging player for the central action of AMP-activated protein kinase. *Biochem Biophys Res Commun* 421, 1-3.

Darby, M.M., Serebreni, L., Pan, X., Boeke, J.D., and Corden, J.L. (2012). The *Saccharomyces cerevisiae* Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion. *Mol Cell Biol* 32, 1762-1775.

de Mestier, L., Pasmant, E., Fleury, C., Brixi, H., Sohier, P., Feron, T., Diebold, M.D., Clauser, E., Cadiot, G., and Groupe d'Etude des Tumeurs, E. (2017). Familial small-intestine carcinoids: Chromosomal alterations and germline inositol polyphosphate multikinase sequencing. *Dig Liver Dis* 49, 98-102.

Desmyter, A., Spinelli, S., Roussel, A., and Cambillau, C. (2015). Camelid nanobodies: killing two birds with one stone. *Curr Opin Struct Biol* 32, 1-8.

Dewaste, V., Moreau, C., De Smedt, F., Bex, F., De Smedt, H., Wuytack, F., Missiaen, L., and Erneux, C. (2003). The three isoenzymes of human inositol-1,4,5-trisphosphate 3-kinase show specific intracellular localization but comparable Ca²⁺ responses on transfection in COS-7 cells. *Biochem J* 374, 41-49.

Dewaste, V., Pouillon, V., Moreau, C., Shears, S., Takazawa, K., and Erneux, C. (2000). Cloning and expression of a cDNA encoding human inositol 1,4,5-trisphosphate 3-kinase C. *Biochem J* 352 Pt 2, 343-351.

Draskovic, P., Saiardi, A., Bhandari, R., Burton, A., Ilc, G., Kovacevic, M., Snyder, S.H., and Podobnik, M. (2008). Inositol hexakisphosphate kinase products contain diphosphate and triphosphate groups. *Chem Biol* 15, 274-286.

Duhoo, Y., Roche, J., Trinh, T.T.N., Desmyter, A., Gaubert, A., Kellenberger, C., Cambillau, C., Roussel, A., and Leone, P. (2017). Camelid nanobodies used as crystallization chaperones for different constructs of PorM, a component of the type IX secretion system from *Porphyromonas gingivalis*. *Acta Crystallogr F Struct Biol Commun* 73, 286-293.

Efanov, A.M., Zaitsev, S.V., and Berggren, P.O. (1997). Inositol hexakisphosphate stimulates non-Ca²⁺-mediated and primes Ca²⁺-mediated exocytosis of insulin by activation of protein kinase C. *Proc Natl Acad Sci U S A* 94, 4435-4439.

Ehrnstorfer, I.A., Geertsma, E.R., Pardon, E., Steyaert, J., and Dutzler, R. (2014). Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. *Nat Struct Mol Biol* 21, 990-996.

Endo-Streeter, S., Tsui, M.K., Odom, A.R., Block, J., and York, J.D. (2012). Structural studies and protein engineering of inositol phosphate multikinase. *J Biol Chem* 287, 35360-35369.

Franco-Echevarria, E., Banos-Sanz, J.I., Monterroso, B., Round, A., Sanz-Aparicio, J., and Gonzalez, B. (2014). A new calmodulin-binding motif for inositol 1,4,5-trisphosphate 3-kinase regulation. *Biochem J* 463, 319-328.

Fu, C., Xu, J., Cheng, W., Rojas, T., Chin, A.C., Snowman, A.M., Harraz, M.M., and Snyder, S.H. (2017). Neuronal migration is mediated by inositol hexakisphosphate kinase 1 via alpha-actinin and focal adhesion kinase. *Proc Natl Acad Sci U S A* 114, 2036-2041.

Fu, C., Xu, J., Li, R.J., Crawford, J.A., Khan, A.B., Ma, T.M., Cha, J.Y., Snowman, A.M., Pletnikov, M.V., and Snyder, S.H. (2015). Inositol Hexakisphosphate Kinase-3 Regulates the Morphology and Synapse Formation of Cerebellar Purkinje Cells via Spectrin/Adducin. *J Neurosci* 35, 11056-11067.

Garman, E.F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* 66, 339-351.

- Garman, E.F., and Owen, R.L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 62, 32-47.
- Gonzalez, B., Banos-Sanz, J.I., Villate, M., Brearley, C.A., and Sanz-Aparicio, J. (2010). Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition. *Proc Natl Acad Sci U S A* 107, 9608-9613.
- Gonzalez, B., Schell, M.J., Letcher, A.J., Veprintsev, D.B., Irvine, R.F., and Williams, R.L. (2004). Structure of a human inositol 1,4,5-trisphosphate 3-kinase: substrate binding reveals why it is not a phosphoinositide 3-kinase. *Mol Cell* 15, 689-701.
- Gosein, V., Leung, T.F., Krajden, O., and Miller, G.J. (2012). Inositol phosphate-induced stabilization of inositol 1,3,4,5,6-pentakisphosphate 2-kinase and its role in substrate specificity. *Protein Sci* 21, 737-742.
- Gosein, V., and Miller, G.J. (2013a). Conformational stability of inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) dictates its substrate selectivity. *J Biol Chem* 288, 36788-36795.
- Gosein, V., and Miller, G.J. (2013b). Roles of phosphate recognition in inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) substrate binding and activation. *J Biol Chem* 288, 26908-26913.
- Graf, E., Empson, K.L., and Eaton, J.W. (1987). Phytic acid. A natural antioxidant. *J Biol Chem* 262, 11647-11650.
- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B., Bendahman, N., and Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363, 446-448.
- Hassaine, G., Deluz, C., Grasso, L., Wyss, R., Tol, M.B., Hovius, R., Graff, A., Stahlberg, H., Tomizaki, T., Desmyter, A., *et al.* (2014). X-ray structure of the mouse serotonin 5-HT₃ receptor. *Nature* 512, 276-281.
- Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254, 51-58.
- Ho, M.W., Yang, X., Carew, M.A., Zhang, T., Hua, L., Kwon, Y.U., Chung, S.K., Adelt, S., Vogel, G., Riley, A.M., *et al.* (2002). Regulation of Ins(3,4,5,6)P(4) signaling by a reversible kinase/phosphatase. *Curr Biol* 12, 477-482.
- Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* 286, 3645-3657.
- Holmes, W., and Jogl, G. (2006). Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity. *J Biol Chem* 281, 38109-38116.
- Hoy, M., Efanov, A.M., Bertorello, A.M., Zaitsev, S.V., Olsen, H.L., Bokvist, K., Leibiger, B., Leibiger, I.B., Zwiller, J., Berggren, P.O., *et al.* (2002). Inositol hexakisphosphate promotes dynamin I-mediated endocytosis. *Proc Natl Acad Sci U S A* 99, 6773-6777.
- Hurrell, R.F. (2003). Influence of vegetable protein sources on trace element and mineral bioavailability. *J Nutr* 133, 2973S-2977S.
- Illies, C., Gromada, J., Fiume, R., Leibiger, B., Yu, J., Juhl, K., Yang, S.N., Barma, D.K., Falck, J.R., Saiardi, A., *et al.* (2007). Requirement of inositol pyrophosphates for full exocytotic capacity in pancreatic beta cells. *Science* 318, 1299-1302.
- Irvine, R.F., and Schell, M.J. (2001). Back in the water: the return of the inositol phosphates. *Nat Rev Mol Cell Biol* 2, 327-338.

- Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P.D., Curry, S., and Conte, M.R. (2003). Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. *Structure* *11*, 833-843.
- Jadav, R.S., Chanduri, M.V., Sengupta, S., and Bhandari, R. (2013). Inositol pyrophosphate synthesis by inositol hexakisphosphate kinase 1 is required for homologous recombination repair. *J Biol Chem* *288*, 3312-3321.
- Jamonnak, N., Creamer, T.J., Darby, M.M., Schaughency, P., Wheelan, S.J., and Corden, J.L. (2011). Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* *17*, 2011-2025.
- Kato, H., Uzawa, K., Onda, T., Kato, Y., Saito, K., Nakashima, D., Ogawara, K., Bukawa, H., Yokoe, H., and Tanzawa, H. (2006). Down-regulation of 1D-myo-inositol 1,4,5-trisphosphate 3-kinase A protein expression in oral squamous cell carcinoma. *Int J Oncol* *28*, 873-881.
- Kato, T., Kajikawa, M., Maenaka, K., and Park, E.Y. (2010). Silkworm expression system as a platform technology in life science. *Appl Microbiol Biotechnol* *85*, 459-470.
- Kim, E., Ahn, H., Kim, M.G., Lee, H., and Kim, S. (2017). The Expanding Significance of Inositol Polyphosphate Multikinase as a Signaling Hub. *Mol Cells* *40*, 315-321.
- Kim, E., Beon, J., Lee, S., Park, J., and Kim, S. (2016). IPMK: A versatile regulator of nuclear signaling events. *Adv Biol Regul* *61*, 25-32.
- Kinebuchi, T., Kagawa, W., Enomoto, R., Tanaka, K., Miyagawa, K., Shibata, T., Kurumizaka, H., and Yokoyama, S. (2004). Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol Cell* *14*, 363-374.
- Koldobskiy, M.A., Chakraborty, A., Werner, J.K., Jr., Snowman, A.M., Juluri, K.R., Vandiver, M.S., Kim, S., Heletz, S., and Snyder, S.H. (2010). p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. *Proc Natl Acad Sci U S A* *107*, 20947-20951.
- Koldobskiy, M.A., and Snyder, S.H. (2011). Inositol pyrophosphates in cell death and life. *Cell Cycle* *10*, 568-570.
- Kost, T.A., and Condreay, J.P. (1999). Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr Opin Biotechnol* *10*, 428-433.
- Kost, T.A., Condreay, J.P., and Jarvis, D.L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotechnol* *23*, 567-575.
- Koster, J.D., Leggewie, B., Blechner, C., Brandt, N., Fester, L., Rune, G., Schweizer, M., Kindler, S., and Windhorst, S. (2016). Inositol-1,4,5-trisphosphate-3-kinase-A controls morphology of hippocampal dendritic spines. *Cell Signal* *28*, 83-90.
- Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* *26*, 1891-1896.
- Lee, Y.S., Mulugu, S., York, J.D., and O'Shea, E.K. (2007). Regulation of a cyclin-CDK-CDK inhibitor complex by inositol pyrophosphates. *Science* *316*, 109-112.
- Leonaite, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *EMBO J* *36*, 1590-1604.
- Lin, A.N., Barnes, S., and Wallace, R.W. (1990). Phosphorylation by protein kinase C inactivates an inositol 1,4,5-trisphosphate 3-kinase purified from human platelets. *Biochem Biophys Res Commun* *170*, 1371-1376.

- Lunde, B.M., Horner, M., and Meinhart, A. (2011). Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res* 39, 337-346.
- Luo, H.R., Huang, Y.E., Chen, J.C., Saiardi, A., Iijima, M., Ye, K., Huang, Y., Nagata, E., Devreotes, P., and Snyder, S.H. (2003). Inositol pyrophosphates mediate chemotaxis in *Dictyostelium* via pleckstrin homology domain-PtdIns(3,4,5)P₃ interactions. *Cell* 114, 559-572.
- Luo, H.R., Saiardi, A., Nagata, E., Ye, K., Yu, H., Jung, T.S., Luo, X., Jain, S., Sawa, A., and Snyder, S.H. (2001). GRAB: a physiologic guanine nucleotide exchange factor for Rab3A, which interacts with inositol hexakisphosphate kinase. *Neuron* 31, 439-451.
- Maag, D., Maxwell, M.J., Hardesty, D.A., Boucher, K.L., Choudhari, N., Hanno, A.G., Ma, J.F., Snowman, A.S., Pietropaoli, J.W., Xu, R., *et al.* (2011). Inositol polyphosphate multikinase is a physiologic PI3-kinase that activates Akt/PKB. *Proc Natl Acad Sci U S A* 108, 1391-1396.
- Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., and Bass, B.L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309, 1534-1539.
- Mailleux, P., Takazawa, K., Erneux, C., and Vanderhaeghen, J.J. (1991). Inositol 1,4,5-trisphosphate 3-kinase distribution in the rat brain. High levels in the hippocampal CA1 pyramidal and cerebellar Purkinje cells suggest its involvement in some memory processes. *Brain Res* 539, 203-210.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298, 1912-1934.
- Manoutcharian, K., Perez-Garmendia, R., and Gevorkian, G. (2017). Recombinant Antibody Fragments for Neurodegenerative Diseases. *Curr Neuroparmacol* 15, 779-788.
- Mansilla, A., Chaves-Sanjuan, A., Campillo, N.E., Semelidou, O., Martinez-Gonzalez, L., Infantes, L., Gonzalez-Rubio, J.M., Gil, C., Conde, S., Skoulakis, E.M., *et al.* (2017). Interference of the complex between NCS-1 and Ric8a with phenothiazines regulates synaptic function and is an approach for fragile X syndrome. *Proc Natl Acad Sci U S A* 114, E999-E1008.
- Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272, 2118-2131.
- Martinez-Lumbreras, S., Taverniti, V., Zorrilla, S., Seraphin, B., and Perez-Canadillas, J.M. (2016). Gbp2 interacts with THO/TREX through a novel type of RRM domain. *Nucleic Acids Res* 44, 437-448.
- Metz, S.W., and Pijlman, G.P. (2011). Arbovirus vaccines; opportunities for the baculovirus-insect cell expression system. *J Invertebr Pathol* 107 Suppl, S16-30.
- Miller, G.J., and Hurley, J.H. (2004). Crystal structure of the catalytic core of inositol 1,4,5-trisphosphate 3-kinase. *Mol Cell* 15, 703-711.
- Miller, G.J., Wilson, M.P., Majerus, P.W., and Hurley, J.H. (2005). Specificity determinants in inositol polyphosphate synthesis: crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase. *Mol Cell* 18, 201-212.
- Misra, S., and Hurley, J.H. (1999). Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell* 97, 657-666.
- Monnier, V., Girardot, F., Audin, W., and Tricoire, H. (2002). Control of oxidative stress resistance by IP3 kinase in *Drosophila melanogaster*. *Free Radic Biol Med* 33, 1250-1259.
- Montpetit, B., Thomsen, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 472, 238-242.

Moritoh, Y., Oka, M., Yasuhara, Y., Hozumi, H., Iwachidow, K., Fuse, H., and Tozawa, R. (2016). Inositol Hexakisphosphate Kinase 3 Regulates Metabolism and Lifespan in Mice. *Sci Rep* 6, 32072.

Morrison, B.H., Bauer, J.A., Kalvakolanu, D.V., and Lindner, D.J. (2001). Inositol hexakisphosphate kinase 2 mediates growth suppressive and apoptotic effects of interferon-beta in ovarian carcinoma cells. *J Biol Chem* 276, 24965-24970.

Morrison, B.H., Bauer, J.A., Lupica, J.A., Tang, Z., Schmidt, H., DiDonato, J.A., and Lindner, D.J. (2007). Effect of inositol hexakisphosphate kinase 2 on transforming growth factor beta-activated kinase 1 and NF-kappaB activation. *J Biol Chem* 282, 15349-15356.

Morrison, B.H., Haney, R., Lamarre, E., Drazba, J., Prestwich, G.D., and Lindner, D.J. (2009). Gene deletion of inositol hexakisphosphate kinase 2 predisposes to aerodigestive tract carcinoma. *Oncogene* 28, 2383-2392.

Mulugu, S., Bai, W., Fridy, P.C., Bastidas, R.J., Otto, J.C., Dollins, D.E., Haystead, T.A., Ribeiro, A.A., and York, J.D. (2007). A conserved family of enzymes that phosphorylate inositol hexakisphosphate. *Science* 316, 106-109.

Murphy, A.M., Otto, B., Brearley, C.A., Carr, J.P., and Hanke, D.E. (2008). A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. *Plant J* 56, 638-652.

Naeini, M.M., and Ardekani, A.M. (2009). Noncoding RNAs and Cancer. *Avicenna J Med Biotechnol* 1, 55-70.

Nagata, E., Luo, H.R., Saiardi, A., Bae, B.I., Suzuki, N., and Snyder, S.H. (2005). Inositol hexakisphosphate kinase-2, a physiologic mediator of cell death. *J Biol Chem* 280, 1634-1640.

Nagata, E., Saiardi, A., Tsukamoto, H., Okada, Y., Itoh, Y., Satoh, T., Itoh, J., Margolis, R.L., Takizawa, S., Sawa, A., *et al.* (2011). Inositol hexakisphosphate kinases induce cell death in Huntington disease. *J Biol Chem* 286, 26680-26686.

Nagata, E., Saiardi, A., Tsukamoto, H., Satoh, T., Itoh, Y., Itoh, J., Shibata, M., Takizawa, S., and Takagi, S. (2010). Inositol hexakisphosphate kinases promote autophagy. *Int J Biochem Cell Biol* 42, 2065-2071.

Nettleship, J.E., Assenberg, R., Diprose, J.M., Rahman-Huq, N., and Owens, R.J. (2010). Recent advances in the production of proteins in insect and mammalian cells for structural biology. *J Struct Biol* 172, 55-65.

Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L., *et al.* (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309, 2054-2057.

Odom, A.R., Stahlberg, A., Wente, S.R., and York, J.D. (2000). A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. *Science* 287, 2026-2029.

Oh, J., Lee, D.G., Bahn, Y.S., and Rhee, S. (2017). Crystal structure of inositol 1,3,4,5,6-pentakisphosphate 2-kinase from *Cryptococcus neoformans*. *J Struct Biol*.

Pacold, M.E., Suire, S., Perisic, O., Lara-Gonzalez, S., Davis, C.T., Walker, E.H., Hawkins, P.T., Stephens, L., Eccleston, J.F., and Williams, R.L. (2000). Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. *Cell* 103, 931-943.

Padmanabhan, U., Dollins, D.E., Fridy, P.C., York, J.D., and Downes, C.P. (2009). Characterization of a selective inhibitor of inositol hexakisphosphate kinases: use in defining biological roles and metabolic relationships of inositol pyrophosphates. *J Biol Chem* 284, 10571-10582.

Papke, B., and Der, C.J. (2017). Drugging RAS: Know the enemy. *Science* 355, 1158-1163.

- Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G., Wohlfonig, A., Ruf, A., Muyldermans, S., Hol, W.G., Kobilka, B.K., and Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* 9, 674-693.
- Pathare, G.R., Nagy, I., Sledz, P., Anderson, D.J., Zhou, H.J., Pardon, E., Steyaert, J., Forster, F., Bracher, A., and Baumeister, W. (2014). Crystal structure of the proteasomal deubiquitylation module Rpn8-Rpn11. *Proc Natl Acad Sci U S A* 111, 2984-2989.
- Pattani, K., and Banting, G. (2004). Ins(1,4,5)P3 metabolism and the family of IP3-3Kinases. *Cell Signal* 16, 643-654.
- Perez Canadillas, J.M., and Varani, G. (2003). Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* 22, 2821-2830.
- Phillippy, B.Q., Ullah, A.H., and Ehrlich, K.C. (1994). Purification and some properties of inositol 1,3,4,5,6-Pentakisphosphate 2-kinase from immature soybean seeds. *J Biol Chem* 269, 28393-28399.
- Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J* 31, 3935-3948.
- Pouillon, V., Marechal, Y., Fripiat, C., Erneux, C., and Schurmans, S. (2013). Inositol 1,4,5-trisphosphate 3-kinase B (Itpkb) controls survival, proliferation and cytokine production in mouse peripheral T cells. *Adv Biol Regul* 53, 39-50.
- Quan, Z., Zheng, D., and Qing, H. (2017). Regulatory Roles of Long Non-Coding RNAs in the Central Nervous System and Associated Neurodegenerative Diseases. *Front Cell Neurosci* 11, 175.
- Quinlan, L.R. (2006). Phosphoinositides, inositol phosphates, and phospholipase C in embryonic stem cells. *Methods Mol Biol* 329, 127-149.
- Raboy, V. (2003). myo-Inositol-1,2,3,4,5,6-hexakisphosphate. *Phytochemistry* 64, 1033-1043.
- Ravelli, R.B., and Garman, E.F. (2006). Radiation damage in macromolecular cryocrystallography. *Curr Opin Struct Biol* 16, 624-629.
- Resnick, A.C., and Saiardi, A. (2008). Inositol polyphosphate multikinase: metabolic architect of nuclear inositides. *Front Biosci* 13, 856-866.
- Resnick, A.C., Snowman, A.M., Kang, B.N., Hurt, K.J., Snyder, S.H., and Saiardi, A. (2005). Inositol polyphosphate multikinase is a nuclear PI3-kinase with transcriptional regulatory activity. *Proc Natl Acad Sci U S A* 102, 12783-12788.
- Rosser, M.P., Xia, W., Hartsell, S., McCaman, M., Zhu, Y., Wang, S., Harvey, S., Bringmann, P., and Cobb, R.R. (2005). Transient transfection of CHO-K1-S using serum-free medium in suspension: a rapid mammalian protein expression system. *Protein Expr Purif* 40, 237-243.
- Rossmann, M.G. (2001). Molecular replacement--historical background. *Acta Crystallogr D Biol Crystallogr* 57, 1360-1366.
- Rychlowska, M., Gromadzka, B., Bienkowska-Szewczyk, K., and Szewczyk, B. (2011). Application of baculovirus-insect cell expression system for human therapy. *Curr Pharm Biotechnol* 12, 1840-1849.
- Safrany, S.T., Caffrey, J.J., Yang, X., and Shears, S.B. (1999). Diphosphoinositol polyphosphates: the final frontier for inositide research? *Biol Chem* 380, 945-951.
- Saiardi, A. (2012). Cell signalling by inositol pyrophosphates. *Subcell Biochem* 59, 413-443.
- Saiardi, A., Caffrey, J.J., Snyder, S.H., and Shears, S.B. (2000). The inositol hexakisphosphate kinase family. Catalytic flexibility and function in yeast vacuole biogenesis. *J Biol Chem* 275, 24686-24692.
- Saiardi, A., Nagata, E., Luo, H.R., Snowman, A.M., and Snyder, S.H. (2001). Identification and characterization of a novel inositol hexakisphosphate kinase. *J Biol Chem* 276, 39179-39185.

Saiardi, A., Resnick, A.C., Snowman, A.M., Wendland, B., and Snyder, S.H. (2005). Inositol pyrophosphates regulate cell death and telomere length through phosphoinositide 3-kinase-related protein kinases. *Proc Natl Acad Sci U S A* *102*, 1911-1914.

Sarmah, B., Latimer, A.J., Appel, B., and Wente, S.R. (2005). Inositol polyphosphates regulate zebrafish left-right asymmetry. *Developmental cell* *9*, 133-145.

Scapin, G. (2013). Molecular replacement then and now. *Acta Crystallogr D Biol Crystallogr* *69*, 2266-2275.

Schell, M.J. (2010). Inositol trisphosphate 3-kinases: focus on immune and neuronal signaling. *Cell Mol Life Sci* *67*, 1755-1778.

Schell, M.J., Erneux, C., and Irvine, R.F. (2001). Inositol 1,4,5-trisphosphate 3-kinase A associates with F-actin and dendritic spines via its N terminus. *J Biol Chem* *276*, 37537-37546.

Scherer, P.C., Ding, Y., Liu, Z., Xu, J., Mao, H., Barrow, J.C., Wei, N., Zheng, N., Snyder, S.H., and Rao, F. (2016). Inositol hexakisphosphate (IP6) generated by IP5K mediates cullin-COP9 signalosome interactions and CRL function. *Proc Natl Acad Sci U S A* *113*, 3503-3508.

Schroder, D., Todter, K., Gonzalez, B., Franco-Echevarria, E., Rohaly, G., Blecher, C., Lin, H.Y., Mayr, G.W., and Windhorst, S. (2015). The new InsP3Kinase inhibitor BIP-4 is competitive to InsP3 and blocks proliferation and adhesion of lung cancer cells. *Biochem Pharmacol* *96*, 143-150.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* *155*, 1075-1087.

Sei, Y., Zhao, X., Forbes, J., Szymczak, S., Li, Q., Trivedi, A., Voellinger, M., Joy, G., Feng, J., Whatley, M., *et al.* (2015). A Hereditary Form of Small Intestinal Carcinoid Associated With a Germline Mutation in Inositol Polyphosphate Multikinase. *Gastroenterology* *149*, 67-78.

Seo, M.D., Velamakanni, S., Ishiyama, N., Stathopoulos, P.B., Rossi, A.M., Khan, S.A., Dale, P., Li, C., Ames, J.B., Ikura, M., *et al.* (2012). Structural and functional conservation of key domains in InsP3 and ryanodine receptors. *Nature* *483*, 108-112.

Shames, D.S., and Minna, J.D. (2008). IP6K2 is a client for HSP90 and a target for cancer therapeutics development. *Proc Natl Acad Sci U S A* *105*, 1389-1390.

Shears, S.B., Ganapathi, S.B., Gokhale, N.A., Schenk, T.M., Wang, H., Weaver, J.D., Zaremba, A., and Zhou, Y. (2012). Defining signal transduction by inositol phosphates. *Subcell Biochem* *59*, 389-412.

Shen, X., Xiao, H., Ranallo, R., Wu, W.H., and Wu, C. (2003). Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science* *299*, 112-114.

Siegemund, S., Rigaud, S., Conche, C., Broaten, B., Schaffer, L., Westernberg, L., Head, S.R., and Sauer, K. (2015). IP3 3-kinase B controls hematopoietic stem cell homeostasis and prevents lethal hematopoietic failure in mice. *Blood* *125*, 2786-2797.

Solyakov, L., Cain, K., Tracey, B.M., Jukes, R., Riley, A.M., Potter, B.V., and Tobin, A.B. (2004). Regulation of casein kinase-2 (CK2) activity by inositol phosphates. *J Biol Chem* *279*, 43403-43410.

Stephens, L.R., Hawkins, P.T., Stanley, A.F., Moore, T., Poyner, D.R., Morris, P.J., Hanley, M.R., Kay, R.R., and Irvine, R.F. (1991). myo-inositol pentakisphosphates. Structure, biological occurrence and phosphorylation to myo-inositol hexakisphosphate. *Biochem J* *275* (Pt 2), 485-499.

Strub, M.P., Hoh, F., Sanchez, J.F., Strub, J.M., Bock, A., Aumelas, A., and Dumas, C. (2003). Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* *11*, 1359-1367.

Stygelbout, V., Leroy, K., Pouillon, V., Ando, K., D'Amico, E., Jia, Y., Luo, H.R., Duyckaerts, C., Erneux, C., Schurmans, S., *et al.* (2014). Inositol trisphosphate 3-kinase B is increased in human Alzheimer brain and exacerbates mouse Alzheimer pathology. *Brain* 137, 537-552.

Szijgyarto, Z., Garedew, A., Azevedo, C., and Saiardi, A. (2011). Influence of inositol pyrophosphates on cellular energy dynamics. *Science* 334, 802-805.

Thota, S.G., and Bhandari, R. (2015). The emerging roles of inositol pyrophosphates in eukaryotic cell physiology. *J Biosci* 40, 593-605.

Thota, S.G., Unnikannan, C.P., Thampatty, S.R., Manorama, R., and Bhandari, R. (2015). Inositol pyrophosphates regulate RNA polymerase I-mediated rRNA transcription in *Saccharomyces cerevisiae*. *Biochem J* 466, 105-114.

Tsubokawa, H., Oguro, K., Robinson, H.P., Masuzawa, T., Rhee, T.S., Takenawa, T., and Kawai, N. (1994). Inositol 1,3,4,5-tetrakisphosphate as a mediator of neuronal death in ischemic hippocampus. *Neuroscience* 59, 291-297.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., *et al.* (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell* 55, 467-481.

Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* 21, 239-248.

Verbsky, J., Lavine, K., and Majerus, P.W. (2005). Disruption of the mouse inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene, associated lethality, and tissue distribution of 2-kinase expression. *Proc Natl Acad Sci U S A* 102, 8448-8453.

Verbsky, J.W., Wilson, M.P., Kisseleva, M.V., Majerus, P.W., and Wente, S.R. (2002). The synthesis of inositol hexakisphosphate. Characterization of human inositol 1,3,4,5,6-pentakisphosphate 2-kinase. *J Biol Chem* 277, 31857-31862.

Walker, E.H., Perisic, O., Ried, C., Stephens, L., and Williams, R.L. (1999). Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* 402, 313-320.

Wang, H., DeRose, E.F., London, R.E., and Shears, S.B. (2014a). IP6K structure and the molecular determinants of catalytic specificity in an inositol phosphate kinase family. *Nat Commun* 5, 4178.

Wang, H., Falck, J.R., Hall, T.M., and Shears, S.B. (2011). Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding. *Nat Chem Biol* 8, 111-116.

Wang, H., Godage, H.Y., Riley, A.M., Weaver, J.D., Shears, S.B., and Potter, B.V. (2014b). Synthetic inositol phosphate analogs reveal that PPIP5K2 has a surface-mounted substrate capture site that is a target for drug discovery. *Chem Biol* 21, 689-699.

Wang, H., and Shears, S.B. (2017). Structural features of human inositol phosphate multikinase rationalize its inositol phosphate kinase and phosphoinositide 3-kinase activities. *J Biol Chem*.

Wickramasinghe, V.O., Savill, J.M., Chavali, S., Jonsdottir, A.B., Rajendra, E., Gruner, T., Laskey, R.A., Babu, M.M., and Venkitaraman, A.R. (2013). Human inositol polyphosphate multikinase regulates transcript-selective nuclear mRNA export to preserve genome integrity. *Mol Cell* 51, 737-750.

Wilson, M.A., Meaux, S., and van Hoof, A. (2007). A genomic screen in yeast reveals novel aspects of nonstop mRNA metabolism. *Genetics* 177, 773-784.

Wilson, M.S., Livermore, T.M., and Saiardi, A. (2013). Inositol pyrophosphates: between signalling and metabolism. *Biochem J* 452, 369-379.

- Windhorst, S., Kalinina, T., Schmid, K., Blechner, C., Kriebitzsch, N., Hinsch, R., Chang, L., Herich, L., Schumacher, U., and Mayr, G.W. (2011). Functional role of inositol-1,4,5-trisphosphate-3-kinase-A for motility of malignant transformed cells. *Int J Cancer* 129, 1300-1309.
- Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.H., Kecman, T., *et al.* (2017). The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat Commun* 8, 14861.
- Wu, Y., Hao, X., and Li, M. (2017). [Application of nanobody in cancer treatment]. *Sheng Wu Gong Cheng Xue Bao* 33, 1085-1090.
- Xu, J., Brearley, C.A., Lin, W.H., Wang, Y., Ye, R., Mueller-Roeber, B., Xu, Z.H., and Xue, H.W. (2005). A role of Arabidopsis inositol polyphosphate kinase, AtIPK2alpha, in pollen germination and root growth. *Plant Physiol* 137, 94-103.
- Xu, R., Sen, N., Paul, B.D., Snowman, A.M., Rao, F., Vandiver, M.S., Xu, J., and Snyder, S.H. (2013). Inositol polyphosphate multikinase is a coactivator of p53-mediated transcription and cell death. *Sci Signal* 6, ra22.
- Yamamura, S., Imai-Sumida, M., Tanaka, Y., and Dahiya, R. (2017). Interaction and cross-talk between non-coding RNAs. *Cell Mol Life Sci*.
- Yang, S.N., Yu, J., Mayr, G.W., Hofmann, F., Larsson, O., and Berggren, P.O. (2001). Inositol hexakisphosphate increases L-type Ca²⁺ channel activity by stimulation of adenylyl cyclase. *FASEB J* 15, 1753-1763.
- York, J.D., Odom, A.R., Murphy, R., Ives, E.B., and Went, S.R. (1999). A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 285, 96-100.
- York, S.J., Armbruster, B.N., Greenwell, P., Petes, T.D., and York, J.D. (2005). Inositol diphosphate signaling regulates telomere length. *J Biol Chem* 280, 4264-4269.
- Yu, W., Ye, C., and Greenberg, M.L. (2016). Inositol Hexakisphosphate Kinase 1 (IP6K1) Regulates Inositol Synthesis in Mammalian Cells. *J Biol Chem* 291, 10437-10444.

8. ANEXO: MATERIALES Y MÉTODOS

8. Anexo: Materiales y métodos

8.1 Introducción a la biología estructural

La Biología Estructural es una rama de la ciencia cuyo objetivo es el estudio de la estructura de macromoléculas biológicas tales como las proteínas y ácidos nucleicos y su relación con la función biológica. Para lograr determinar la forma tridimensional de las macromoléculas biológicas es necesario el uso combinado de técnicas de Biología Molecular, Bioquímica y Biofísica. El estudio estructural es necesario para llegar a discernir los mecanismos de acción de las proteínas y el ADN o el ARN en la célula. El conocimiento de la estructura tridimensional de una macromolécula nos permite entender su función, por lo que hablamos de una relación estructura-función. Esta estructura no solo arroja información sobre los procesos biológicos, sino que también tiene importantes aplicaciones prácticas como son la ingeniería de proteínas o el diseño de fármacos.

Entre las técnicas empleadas para el análisis estructural de las biomoléculas destacan la Cristalografía de rayos X, la Microscopía electrónica y la Resonancia magnética nuclear (RMN). La Cristalografía de rayos X permite determinar la estructura de proteínas sin límite en su masa molecular lo que supone una ventaja importante, siendo una de las técnicas experimentales más potentes para determinar la estructura de las macromoléculas biológicas a escala atómica. Mediante esta técnica se han resuelto estructuras tanto de proteínas sencillas como complejos macromoleculares o incluso virus (Figura 8.1).

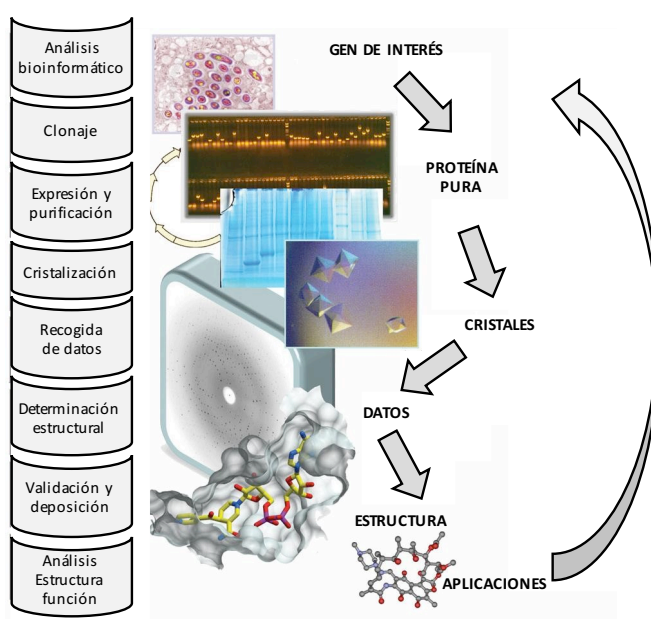


Figura 8.1. Etapas de la resolución estructural de proteínas mediante Cristalografía de rayos X.

La resolución estructural de proteínas mediante la Cristalografía de rayos X implica distintas etapas experimentales dirigidas a la obtención de la estructura y su análisis (resumidas en la figura 8.1). Ésta ha sido la técnica que se ha empleado a lo largo de la presente tesis para determinar la estructura tridimensional de las proteínas que se han estudiado.

8.2. Obtención de la proteína

El primer paso de la resolución de estructuras mediante Cristalografía de rayos X es obtener una muestra de proteína adecuada para su cristalización. Como se verá más adelante esta muestra tiene que cumplir una serie de requisitos que aumenten las probabilidades de obtener cristales (ver sección 8.2.3.2). Los pasos a seguir para la obtención de la muestra son la preparación de construcciones plasmídicas, la expresión y la purificación. Cada proteína seguirá un proceso exclusivo de obtención en base a sus propiedades físico-químicas y biológicas.

8.2.1 Clonación y expresión

La clonación es el proceso por el cual el cDNA (ADN complementario sintetizado a partir del ARNm) codificante de nuestra proteína se introduce en vectores que permitan la expresión de la proteína en los organismos seleccionados. La elección del vector va ligado a la elección del organismo en el que se producirá la proteína y es clave para la obtención de una cantidad adecuada de proteína. El vector permite también introducir diversas etiquetas que faciliten la purificación de la proteína una vez expresada. En ocasiones es necesario modificar determinados elementos en el cDNA truncándolo o realizando mutaciones puntuales en su secuencia (mutagénesis dirigida) con distintos objetivos: favorecer la producción de la proteína de interés, inactivar una enzima y capturar el complejo en presencia de su sustrato, modificar la solubilidad de regiones expuestas al solvente, evitar la flexibilidad de dominios o incluso crear puentes de unión entre diferentes proteínas que forman un complejo.

El vector utilizado mayoritariamente en esta tesis ha sido pKLSLt. Este vector deriva del vector comercial pET-28a (+) (Novagen), por lo que presenta el promotor de la polimerasa del fago T7, resistencia a kanamicina y la capacidad de expresar el represor del operón lactosa (*lacI*), que se inhibe bajo la presencia de IPTG. En ausencia de IPTG el producto del gen *lacI* se une al operador de *lacUV5*, lo que impide la transcripción de la ARN polimerasa. Cuando se induce la expresión con IPTG, este compuesto se une al represor *lacI* impidiendo que se una al operador y permitiendo así la expresión de la ARN polimerasa y de la proteína recombinante codificada en el vector de expresión bajo el control del promotor de T7. Este vector introduce la etiqueta LSLt en el extremo N-terminal, la cual contiene expresión de una lectina, seguida de un sitio de reconocimiento por la endoproteasa del virus del mosaico del tabaco (TEV), cuya secuencia consenso es ENLYFQG. De esta manera se puede eliminar la lectina mediante una digestión controlada con la proteasa TEV (Angulo et al., 2011).

Una vez obtenido el vector de expresión, se introduce en el organismo seleccionado para proceder a la expresión de la proteína. La elección del sistema de expresión está marcada por los requerimientos de la proteína a expresar. Los organismos más empleados para la síntesis de proteínas recombinantes suelen ser bacterias como *Escherichia coli*. Existen multitud de cepas diferentes y cada una posee unas características únicas, la elección de la cepa bacteriana es fundamental para conseguir unos buenos resultados.

8.2.1.1 Expresión en células de insecto

Aunque las diversas tecnologías para la expresión de proteínas recombinantes hayan avanzado considerablemente, en muchos casos, la producción de proteínas sigue siendo un reto. La obtención de

cantidad suficiente de proteína homogénea y soluble representa una etapa limitante en la Cristalografía de proteínas. El empleo de cepas bacterias como método de expresión es ampliamente utilizado debido a su fácil manejo y a su bajo coste. Por ello, actualmente existen un gran número de plásmidos de expresión y cepas especializadas, sin embargo, este sistema presenta una serie de limitaciones para la síntesis de proteínas eucariotas. En particular, las bacterias son incapaces de llevar a cabo modificaciones postraduccionales y en ocasiones no son capaces de generar un plegamiento correcto de la proteína (Baneyx, 1999; Brondyk, 2009).

La creciente tendencia hacia el estudio de proteínas biológicas complejas ha requerido el desarrollo de sistemas de expresión eucariotas más eficaces como el sistema de expresión de baculovirus (BEVs). Este sistema es particularmente útil para producir proteínas complejas, así como grandes complejos de proteínas eucariotas en la cantidad y calidad requeridas para un estudio estructural detallado. Hoy en día, multitud de proteínas recombinantes, tanto citosólicas como proteínas de membrana, se han producido con éxito en células de insecto infectadas con baculovirus. Así, en los últimos 20 años, el sistema de expresión de baculovirus se ha convertido en uno de los sistemas más utilizados para la producción rutinaria de proteínas recombinantes (Kost y Condreay, 1999).

Las células de insecto infectadas por baculovirus son capaces de sobreexpresar las proteínas de interés con altos rendimientos. Lo más importante del sistema de expresión de baculovirus es que promueve el correcto plegamiento de proteínas, lo que se debe a que las células de insecto son eucariotas y por tanto capaces de realizar un plegamiento de proteínas que se asemeja estrechamente al de las células de mamífero. Además, las células de insecto llevan a cabo muchas modificaciones postraduccionales como la glicosilación, palmitolación, miristolación, acilación de ácidos grasos y amidación presentes en células de mamífero. Estas dos características son en muchos casos cruciales para la adecuada función de la proteína (Kato et al., 2010; Kost et al., 2005; Rosser et al., 2005). La principal desventaja de este sistema son los elevados costes en comparación con los sistemas de expresión bacterianos y de levaduras, tanto en materiales como en equipamiento (Assenberg et al., 2013; Nettleship et al., 2010). Además de ser actualmente una de las principales plataformas para la producción de proteínas recombinantes para su posterior estudio estructural, los baculovirus son utilizados para el desarrollo de vacunas (Cox, 2012; Metz y Pijlman, 2011; Rychlowska et al., 2011).

Los baculovirus son virus de ADN de doble cadena que infectan a un gran número de especies de insectos. El baculovirus más utilizado para los estudios de expresión es el virus de poliedrosis nuclear múltiple de *Autographa californica* (AcMNPV) (Ayres et al., 1994), que infecta a las especies de lepidópteros *Spodoptera frugiperda* como insectos huéspedes. Las células SF21 y SF9 (que derivan de las anteriores) se utilizan tradicionalmente para los sistemas de expresión basados en baculovirus y provienen del tejido ovárico de larvas del lepidóptero. En el sistema de “*Bac-to-Bac*” el plásmido donador contiene el gen de interés de nuestra proteína insertado. Este plásmido se utiliza para transformar células competentes que contienen el bácmido (que a su vez contiene el genoma del baculovirus) y un plásmido auxiliar o “*helper*” que codifica para la enzima responsable de la transposición, la transposasa. Una vez transformadas las bacterias con el plásmido donador, ocurre la transposición entre éste y el bácmido, y el gen de interés es insertado en el genoma del baculovirus. A continuación, se extrae el bácmido recombinante y se utiliza para transfectar células de insecto y conseguir la producción de partículas virales recombinantes. Estas pueden ser utilizadas para infectar células nuevamente para conseguir la expresión de la proteína de interés. El baculovirus con el gen de

interés toma el control de la maquinaria de expresión génica de la célula huésped, replicándose utilizando la maquinaria metabólica de las células de insecto.

8.2.2 Purificación

Una vez expresada la proteína de interés se inicia un proceso de purificación en el que se captura la proteína de la bacteria, y se somete a una serie de tratamientos de separación de otras macromoléculas hasta alcanzar el mayor grado de pureza posible. En primer lugar, se realiza la extracción del contenido celular mediante la lisis del cultivo, existen diferentes técnicas como el empleo de ultrasonidos (sonicación) o mediante la aplicación de presión (prensa de French). Para conseguir un mejor resultado en la lisis pueden añadirse inhibidores de proteasas, ligandos, detergentes o agentes reductores. Posteriormente se eliminan los restos celulares por ultracentrifugación. Después, se combinan varios pasos en los que se selecciona la proteína de interés mediante diversas técnicas cromatográficas. En muchos casos se comienza con cromatografías de afinidad y se concluye con una cromatografía de exclusión molecular. Los resultados de cada paso de purificación se analizan mediante una electroforesis desnaturizante en gel de poliacrilamida (SDS-PAGE). Es necesario tener en cuenta que, en cada paso de la purificación, se obtiene proteína cada vez más pura, pero también se pierde una gran cantidad de proteína total. A continuación, se detallan los principales tipos de cromatografías:

- La **cromatografía de afinidad** se basa en la afinidad de una proteína concreta por un compuesto unido a una resina, como puede ser un anticuerpo o un ligando. La interacción reversible entre dicho compuesto y la proteína permite la separación de la proteína del resto de compuestos del extracto proteico. La principal ventaja de la técnica es la gran selectividad que ofrece. Para eluir la proteína de interés debe revertirse la interacción con la columna. Dicha elución puede ser específica (añadiendo un ligando que compita con la proteína por la unión a la resina) o inespecífica (aumentando la fuerza iónica o cambiando el pH). En este trabajo se han utilizado columnas de afinidad como la heparina y la resina de *Sefarosa* CL-6B. La resina de heparina tiene la capacidad de retener proteínas que unen compuestos fosforilados, como las proteínas de unión a ADN. Las quinasas objeto de estudio en esta tesis unen inositoles que están cargados de grupos fosfato por lo que quedan muy retenidas en la columna de Heparina. Para eluir después la proteína basta con realizar un gradiente en el que se aumente de manera gradual la concentración de sal. La *Sefarosa* CL-6B es una resina que posee muy alta afinidad por la etiqueta de fusión LSL. En la elución, se añade lactosa a una concentración 0.2M compitiendo con la resina y desplazando la proteína de interés.
- Las **cromatografías de intercambio iónico** separan las proteínas basándose en las diferencias de carga en su superficie, produciéndose una interacción entre grupos polares cargados de la superficie de la proteína con la resina de la columna. En teoría, para que una proteína interaccione de forma notable con la resina, esta debe estar disuelta en un tampón que presente un pH al menos dos puntos por encima o por debajo del punto isoelectrico (pI) de la proteína. Las proteínas con cargas superficiales opuestas a las de la resina empleada son retenidas, mientras que el resto son eluidas. Las proteínas retenidas son posteriormente eluidas gradualmente usando un gradiente creciente de NaCl en el mismo tampón. Existen dos tipos principales de columnas de intercambio iónico: de intercambio aniónico (Q), que retienen

proteínas cargadas negativamente (pH mayor que el pI de la proteína), y de intercambio catiónico (S), que retienen proteínas cargadas positivamente (pH menor que el pI de la proteína).

- La **cromatografía de exclusión molecular o filtración en gel** separa moléculas en virtud de sus diferencias de tamaño. La capacidad de separación reside en el gel cuya matriz consta de un gran número de esferas porosas constituidas por largas cadenas de polímeros unidas entre sí, formando una extensa red por la que se desplaza la muestra. Cada gel se caracteriza por un intervalo de fraccionamiento que depende del tamaño de sus poros. Las moléculas dentro de dicho intervalo presentan una mayor o menor facilidad para difundir al interior de las esferas en función de su tamaño, siendo extraídas fraccionadamente. Cuanto mayor es una molécula, menor es su capacidad para acceder al interior de las esferas del gel y, por tanto, menor es el volumen de líquido que se requiere para extraerla de la columna. Teniendo en cuenta que para las proteínas globulares el tamaño está, en general, relacionado con el peso molecular, este tipo de cromatografía separa las sustancias en función del peso molecular, obteniéndose primero las más pesadas.

8.3 Cristalización de proteínas

8.3.1 Concepto de cristal

Los cristales son el resultado de una agregación controlada de las moléculas en las que estas se disponen de manera ordenada, simétrica y periódica. Se entiende por cristal una red imaginaria formada por una unidad primaria de volumen, denominada celdilla unidad o celdilla elemental que se repite periódicamente por traslación a lo largo de las tres direcciones del espacio. Esta celdilla está definida por tres ángulos (α , β , γ) y tres ejes (a, b, c), y sirve como sistema de coordenadas para referir las posiciones de cada átomo de la estructura. Dentro de la celdilla elemental pueden existir, a su vez, operaciones de repetición que se denominan elementos de simetría, como ejes de rotación, planos de reflexión y planos de deslizamiento. Así, la unidad asimétrica es el conjunto mínimo de moléculas que, tras aplicar los elementos de simetría, representa el contenido de la celdilla unidad. En el caso de las proteínas, existen 65 modalidades distintas de repetición de esta unidad asimétrica, correspondientes a distintos grupos espaciales. De los 230 grupos posibles, las proteínas únicamente pueden cristalizar en aquellos 65 que no presentan planos de reflexión o centros de inversión, debido a su carácter quiral (Figura 8.2).

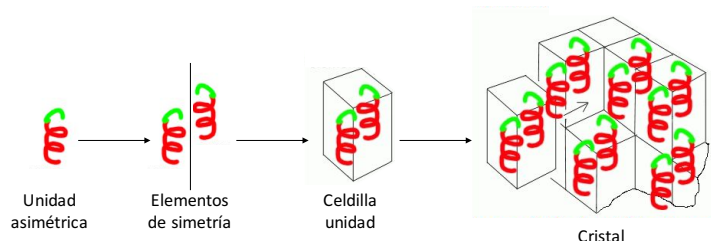


Figura 8.2. Concepto de unidad asimétrica, celdilla unidad y cristal. La unidad asimétrica se repite dentro de la celda unidad. La repetición de la celda unidad genera el cristal. Imagen modificada de: <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

8.3.2 Requisitos previos a la cristalización

Para la resolución de estructuras mediante Cristalografía de rayos X es necesario obtener una muestra de proteína adecuada para su cristalización. Para ello hay que tener en cuenta cinco aspectos fundamentales:

La muestra debe ser **homogénea y pura**. Es indispensable que la muestra presente un alto grado de pureza (90 – 95%). Tanto la pureza como la homogeneidad de la muestra son aspectos críticos que determinan en gran parte el éxito en los experimentos de cristalización. Por norma general se considera adecuada una **concentración** 5-20 mg/ml. Cada proteína necesitará una concentración particular para alcanzar el punto de sobresaturación, por lo que es recomendable realizar un test de precristalización (como la prueba PCT “*Pre-Crystallization Test*”, Hampton Research) que indique si la proteína tiende a precipitar a la concentración elegida. Para muchas proteínas el paso de congelación/descongelación puede ser crítico afectando a su **estabilidad**. Por último, es importante que se pueda producir en **cantidad** suficiente para realizar los ensayos de cristalización y optimizar las condiciones obtenidas. A veces, el uso de diferentes lotes de purificación puede afectar a la cristalización y a la reproducibilidad de los cristales (Bergfors, 2009).

8.3.3 Principios básicos de cristalización

Una vez obtenida la muestra de proteína, se procede a su cristalización, proceso que representa uno de los mayores obstáculos de la Cristalografía. Si se cumplen los requerimientos anteriormente detallados se incrementan las probabilidades de conseguir cristales de proteína. Mediante el diagrama de solubilidad (Figura 8.3) se explica el proceso de formación de cristales.

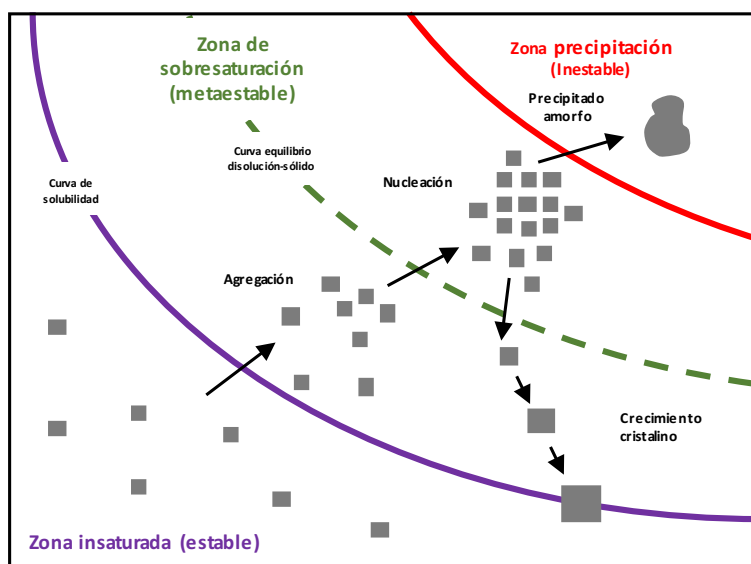


Figura 8.3. Diagrama de solubilidad en el que se muestran las diferentes zonas en las que se puede encontrar la proteína durante un experimento de cristalización. Las flechas indican el comportamiento habitual de la formación de cristales de proteína a partir de una muestra soluble. La zona metaestable se comporta como un gradiente en el que las proteínas se agregan progresivamente hasta formar núcleos.

Se puede dividir el diagrama en cuatro áreas:

- Zona de subsaturación o insaturación donde no se obtienen cristales y la proteína es completamente soluble.
- Área de baja sobresaturación denominada zona metaestable, donde se produce el proceso de crecimiento cristalino.
- Área de moderada sobresaturación, donde se produce el proceso de nucleación.
- Área de sobresaturación denominada zona de precipitación, donde la proteína precipita de forma no controlada y rápida, formando agregados amorfos.

El proceso de cristalización consta de dos etapas, una primera etapa de nucleación y, una vez que el núcleo del cristal alcanza un tamaño crítico se produce la segunda etapa de crecimiento del cristal. Cuando se alcanza la sobresaturación de la muestra, se inicia una precipitación que si se controla lleva a la formación de cristales. Por el contrario, si no se controla puede provocar una precipitación masiva generando un precipitado amorfo en lugar de cristales individuales. En las situaciones en las que se trabaja bajo la curva de solubilidad, la solución no ha alcanzado la saturación. La estrategia consiste llevar a la proteína en solución a la zona de sobresaturación, es decir que sobrepase la curva de solubilidad, pero en una zona cercana para que la precipitación sea ordenada y controlada. De esta forma, se consiguen pocos núcleos de cristalización sobre los que se inicia la etapa de crecimiento del cristal. Finalmente, el sistema se equilibra al llegar a la curva de solubilidad, deteniéndose el crecimiento del cristal. Por tanto, se debe mantener la proteína cerca de la curva de solubilidad, pero en la zona de sobresaturación.

Existen ciertos factores que afectan en la obtención de soluciones sobresaturadas. Estos son la concentración de proteína, la temperatura, la fuerza iónica, el pH, el tampón de la solución en el que se encuentre la muestra, agentes precipitantes, presencia de aditivos o ligandos. Como se ha mencionado, la etapa de cristalización constituye uno de los principales obstáculos de la Cristalografía macromolecular. No hay una técnica de cristalización que asegure el éxito en este proceso, si bien una acertada manipulación de la muestra puede ayudar enormemente a superar esta etapa.

8.3.4 Métodos de cristalización

Existen diversas técnicas de cristalización de proteínas: difusión de vapor, bajo aceite en *microbatch*, diálisis y contradifusión a través de una interfase. En esta tesis se ha utilizado la difusión de vapor para obtener cristales de proteína en todos los casos. Mediante esta técnica, se enfrenta una gota en la que se mezclan volúmenes iguales o distintos de la proteína en solución y un agente precipitante, ambos a una concentración determinada, a un volumen mayor de dicho precipitante (reservorio), sin que la gota y el reservorio estén en contacto y todo ello aislado del exterior. Se denominan agentes precipitantes cualquier tipo de molécula que pueda ayudar a la agregación progresiva de la proteína. Los precipitantes más habituales contienen sal polímeros orgánicos de diferente peso molecular, como el polietilenglicol (PEG). También pueden contener alcoholes o compuestos volátiles, que modifican la constante dieléctrica del solvente, y cuya evaporación altera la solubilidad de la proteína paulatinamente. Bajo estas condiciones, se produce una evaporación de agua desde la gota hacia el precipitante. Así pues, el volumen de la gota irá disminuyendo, mientras que la concentración de proteína va aumentando. El sistema se va acercando lentamente a la saturación y, posteriormente, a la sobresaturación, pasando por la zona de nucleación en la que se generan los primeros núcleos. Finalmente tendrá lugar la etapa de

crecimiento, en la que la proteína libre se une a los núcleos preformados dando lugar a cristales individuales. Por este motivo, las condiciones de cristalización se ajustan en sucesivos experimentos, intentando acercarse lo máximo posible a la zona óptima de nucleación y crecimiento cristalino. Dentro de esta modalidad existen dos montajes diferentes:

- Gota sentada: El precipitante o reservorio se sitúa en un pocillo que rodea una plataforma elevada sobre la que se sitúa la gota con la mezcla de proteína y precipitante. El sistema se aísla del exterior mediante una película adhesiva impermeable que cubre todo el pocillo (Figura 8.4).
- Gota colgante: El precipitante se sitúa en un pocillo sobre el que se coloca un cubreobjetos siliconado, cuya cara interior transporta la gota con la mezcla de proteína y el precipitante. El aislamiento se consigue mediante la utilización de grasa de vacío entre el pocillo y el cubreobjetos (Figura 8.4).

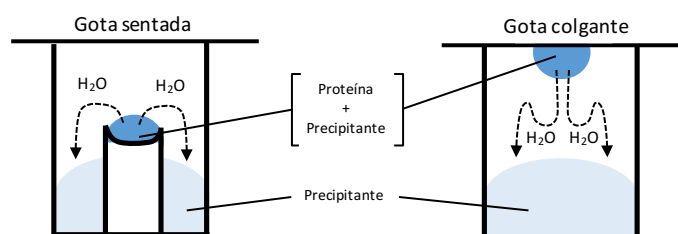


Figura 8.4. Representación esquemática del dispositivo experimental empleado para la cristalización mediante difusión de vapor por gota sentada (izquierda) y gota colgante (derecha).

El proceso de cristalización de cualquier proteína se divide en dos fases principales, una primera fase de búsqueda inicial de condiciones de cristalización y una segunda etapa de optimización.

8.3.5 Búsqueda inicial de condiciones de cristalización

La primera etapa de búsqueda de condiciones de cristalización se realiza mediante técnicas de alto rendimiento que permiten llevar a cabo múltiples experimentos utilizando la mínima cantidad de muestra y tiempo posible. En esta etapa se ha utilizado el robot de cristalización *Innovadine Technologies Inc. NanodropTM ExtY* (Figura 8.5a) y las placas *InnovaplateTM SD-2* (Figura 8.5a), que permiten aplicar la técnica de difusión de vapor en gota sentada y realizar dos experimentos por condición y diversos paquetes comerciales de cristalización como: INDEX y CSIyII (Hampton Research), PACT y JCSG (Qiagen), y JBScreen (Jena Biosciences). Cada placa de cristalización permite investigar 96 condiciones de cristalización diferentes y usar volúmenes de gota muy pequeños (250-500 nL) enfrentadas a un volumen de aproximadamente 65 µl de solución precipitante (Figura 8.5b). Además, se ha utilizado otro robot, el Gilson Quad-Z 215 Liquid Handler que se emplea para preparar las placas con los diferentes kits comerciales (Figura 8.5c).



Figura 8.5. Material utilizado durante la cristalización. a) Robot *Innovadine Technologies Inc. Nanodrop™ ExtY*. b) Placa de 96 pocillos *Innovaplate™ SD-2*. c) Robot *Gilson Quad-Z 215 Liquid Handler*. Imágenes a y c tomadas de <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

8.3.6 Optimización de las condiciones de cristalización

En caso de obtener resultados positivos en una búsqueda inicial, estos frecuentemente corresponden a cristales muy pequeños con irregularidades e imperfecciones. Por ello, se realiza una segunda etapa de optimización cuya duración depende de cada caso en particular y puede ser más o menos laboriosa.

La técnica más usada para la optimización de cristales consiste en variar ligeramente las concentraciones de los componentes de la condición de cristalización. A la vez se suele realizar un escalado, que consiste en aumentar los volúmenes de proteína y precipitante tanto en la gota (hasta 1 o 2 μ l) como en el pocillo (entre 150 y 500 μ l), con el objetivo de conseguir cristales de mayor tamaño. También es habitual realizar una variación de la relación de volúmenes entre la proteína y el precipitante en gota (1:2, 2:1, 1:3, etcétera). Con esto se busca alterar la concentración relativa de la proteína respecto al precipitante. Cambios en el pH y la temperatura, o la inclusión de aditivos, detergentes o ligandos propios de la proteína, son variables a tener en cuenta en esta etapa.

Una técnica que ha sido clave en la obtención de cristales en esta tesis, es la denominada técnica de sembrado o “*seeding*” la cual en muchas ocasiones mejora notablemente la calidad de los cristales (Bergfors, 2003). Esta tiene como objetivo disminuir la sobresaturación del sistema llevándolo a la zona metaestable dentro de la curva de solubilidad, en la cual no puede ocurrir la nucleación espontánea, pero sí puede producirse el crecimiento de cristales a partir de semillas cristalinas. Si bien existen diferentes tipos de sembrado, la base de la técnica es común, consistente en obtener núcleos de un cristal preformado que sirvan como base para el crecimiento de nuevos cristales. En esta tesis se ha trabajado con “soluciones de semillas” (“*seeding stocks*”). Estas soluciones se preparan triturando cristales en la propia condición de cristalización y se pueden congelar para su uso posterior. En concreto, se ha trabajado mucho con la variante de sembrado por arrastre o “*streak-seeding*” la cual ha dado resultados muy favorables (Figura 8.6). Para ello, se introduce una fibra natural en una gota que contiene cristales o bien en una solución de semillas. Esta fibra arrastra núcleos cristalinos que, a continuación, se sumergen en una nueva gota de cristalización. De esta forma se puede realizar un proceso de crecimiento controlado sobre núcleos preformados. Estas técnicas han supuesto un aumento de la reproducibilidad de los cristales, así como una disminución considerable del tiempo necesario para su formación. Además, permite ajustar el número de cristales obtenidos por gota mediante el empleo de diluciones de las soluciones de semillas.

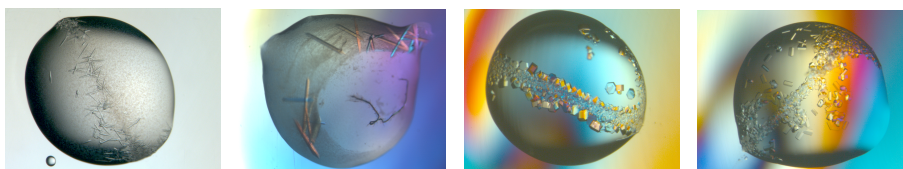


Figura 8.6. Efecto de la técnica del semillado. Ejemplos de gotas de cristalización en las que se ha utilizado la técnica de “*streak-seeding*” en el proceso de optimización de cristales. Se observa el rastro de cristales crecidos por el paso de la fibra natural por la nueva gota.

Por último, la técnica de proteólisis *in situ* facilita en algunos casos la cristalización de proteínas que presentan problemas en la obtención de cristales. Consiste en introducir proteasas en las gotas de cristalización a diferentes concentraciones con el fin de eliminar bucles expuestos que puedan interferir en el proceso de cristalización.

8.3.7 Cristalización de complejos proteína-ligando

Un objetivo muy perseguido en nuestra investigación es la obtención de la estructura tridimensional de la proteína en presencia de algún ligando que interaccione con ella, como puede ser el caso de sustratos o productos de reacción o bien moléculas agonistas/antagonistas de su función. Existen dos estrategias utilizadas para lograr obtener estos complejos: cocrystalización e inmersión.

- Cocrystalización: La proteína se incubaba con el ligando de interés, y después se llevan a cabo los experimentos de cristalización.
- Inmersión, también conocido como “*soaking*”: En esta técnica los cristales de proteína se incuban en una disolución del ligando preferiblemente en la propia condición de cristalización con el fin de evitar un desequilibrio osmótico que deteriore el cristal. Los canales de solvente presentes en los cristales de proteína permiten que el ligando difunda por el cristal y acceda al sitio de unión.

La elección de una aproximación u otra depende de la accesibilidad al sitio de unión al ligando en el cristal, el tamaño del ligando, si existen o no impedimentos estéricos o la estabilidad del cristal en presencia de la solución de ligando. Los tiempos de incubación, así como las concentraciones del ligando son variables. En la mayoría de los casos se prueba con ambas aproximaciones para intentar obtener el complejo proteína-ligando.

8.3.8 Empleo de nanoanticuerpos en la cristalización de proteínas

La producción de cristales y su difracción a alta resolución sigue siendo el principal cuello de botella en la Cristalografía de rayos X. Actualmente está en auge el uso de nanoanticuerpos, “*nanobodies*” o NBs como herramientas auxiliares para cristalizar proteínas. La utilidad de éstos en la biología estructural está firmemente establecida ya que se ha demostrado que actúan a modo de chaperonas en la cristalización de proteínas. Se han logrado grandes resultados mediante el uso de NBs con proteínas flexibles, grandes complejos multiproteicos, estados conformacionales transitorios, proteínas

intrínsecamente desordenadas y, por último, proteínas de membrana (Desmyter et al., 2015; Duhoo et al., 2017; Ehrnstorfer et al., 2014; Hassaine et al., 2014; Pathare et al., 2014).

Los NBs son un tipo de anticuerpos de dominio único presentes en los camélidos. Carecen por completo de cadenas ligeras por lo que consisten sólo en cadenas pesadas (Hamers-Casterman et al., 1993). Son proteínas relativamente sencillas con un tamaño pequeño (14-15 kDa), lo que les proporciona gran estabilidad. Presentan un menor tamaño y mayor estabilidad que los anticuerpos convencionales. Pese a su pequeño tamaño albergan la capacidad de reconocimiento y unión al antígeno. Están codificados por un solo gen, y su naturaleza de dominio simple les permiten ser producidos en gran cantidad usando sistemas de expresión bacteriana. Es preciso señalar que son proteínas extremadamente solubles (40 mg ml⁻¹) y, a menudo se utilizan etiquetas de fusión, como la cola de histidinas, para su purificación. Su estructura contiene uno o dos puentes disulfuro, por lo que generalmente la expresión de los NBs se realiza con vectores que los dirigen al periplasma (Pardon et al., 2014).

Debido a su forma compacta y alargada exhiben una serie de propiedades que pueden ayudar en la cristalización de proteínas. Los NBs tienen acceso a cavidades, hendiduras de la superficie de las proteínas, de esto modo interaccionan con regiones cóncavas y con zonas bisagras de las proteínas facilitando la estabilización de las superficies. Son capaces de insertarse en interfases previniendo la movilidad de ciertos dominios y ayudando también a la estabilización de bucles móviles y grandes complejos. Colaboran con la solubilización de proteínas con baja solubilidad, y proporcionan contactos cristalinos que contribuyen en la formación de cristales de proteína (Desmyter et al., 2015). A su vez presentan resistencia a muchos aditivos, proteasas y a un rango muy amplio de temperatura y pH, por lo que son ideales para el cribado de una amplia gama de condiciones de cristalización con variaciones de pH, fuerza iónica, temperatura, concentración de proteínas, sales, ligandos o aditivos. De este modo las condiciones de cristalización están limitadas en gran medida por la estabilidad de la proteína diana en lugar de por la estabilidad de los NBs.

Sin embargo, a pesar de las múltiples ventajas que presentan, la generación de NBs sigue siendo un procedimiento bastante largo y costoso y, como tal, se limita a casos difíciles de cristalización. Actualmente varias plataformas académicas como INSTRUCT o pequeñas empresas ofrecen servicios que abarcan desde la inmunización de la llama hasta la selección de los NBs. En paralelo, en los últimos años se está estudiando mucho el empleo de NBs en aplicaciones farmacéuticas en el tratamiento contra el cáncer y la enfermedad de Alzheimer por medio de bloqueo de interacciones moleculares (Manoutcharian et al., 2017; Wu et al., 2017).

8.4 Recogida de datos de difracción de rayos X

Una vez superado el cuello de botella al que se enfrenta la Cristalografía y una vez obtenidos cristales de buena calidad, se procede a la toma de datos de difracción que permitirá posteriormente la resolución de su estructura tridimensional. Para ello se hace incidir un haz de rayos X sobre el cristal, el cual mediante el fenómeno de difracción genera múltiples haces que generan un patrón de difracción.

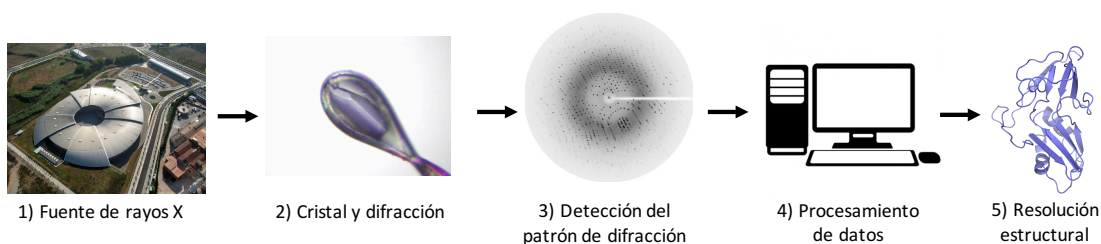


Figura 8.7. Esquema general del experimento de difracción. El experimento de difracción de rayos X se divide en 5 etapas fundamentales: 1) Generación de rayos X. 2) Generación de los haces difractados por el cristal. 3) Detección del patrón de difracción. 4) Análisis y procesamiento de datos. 5) Obtención de la estructura.

Para poder llevar a cabo un experimento de difracción de rayos X con un cristal de proteína se requieren los siguientes elementos:

8.4.1. Elementos de sujeción y protección del cristal

Para realizar una medida estable, los cristales deben estar ubicados en un soporte que los mantenga en una posición fija y que permita su rotación. Además, deben estar protegidos del deterioro que experimentan debido a la exposición a la radiación (Garman, 2010; Garman y Owen, 2006; Ravelli y Garman, 2006). Para reducir este daño al máximo, los cristales se sumergen en nitrógeno líquido para realizar el experimento a una temperatura de 100K (Figura 8.8). Debido al alto contenido de agua de los cristales de proteína, estos han de sumergirse previamente en una solución crioprotectora que desplaza el agua no estructural del cristal evitando la formación de hielo. Entre los crioprotectores más comunes se encuentran el glicerol, etilenglicol, PEGs de diferentes pesos moleculares, criosales o aceites (como paratona o parafina). Durante el proceso, los cristales se capturan de la placa de cristalización en un lazo de nylon de tamaño similar al de cada cristal en particular (Figura 8.8). Posteriormente, en la estación de radiación sincrotrón, los cristales montados en su soporte son transferidos a un goniómetro que permitirá su centrado y rotación durante la recogida de datos a la vez que son sometidos a una corriente constante de nitrógeno cuya temperatura suele situarse en 100 K.



Figura 8.8. Montaje de un cristal en el equipo de difracción. a) Esquema de montaje de un cristal en un bucle con soporte magnético y la posterior inmersión en nitrógeno líquido. b) Arriba, imagen de un cristal montado en el bucle, abajo, estación de radiación sincrotrón. c) Detalle del montaje del cristal para el experimento de difracción. Imagen tomada de <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

8.4.2 Fuentes de rayos X

Los rayos X son radiaciones electromagnéticas cuya longitud de onda se sitúa en el rango de 0,1–100 Å (energías de 123,98 keV a 0,12398 keV). Los rayos X se producen mediante diferentes tipos de fuentes. En este trabajo se han usado dos tipos: instalaciones de radiación sincrotrón y generadores de ánodo rotatorio.

La radiación sincrotrón se produce en grandes instalaciones que integran los siguientes elementos:

- Un acelerador de electrones lineal en el que se inyectan electrones que sufren una aceleración inicial.
- Un anillo de aceleración que eleva su energía hasta valores cercanos a 6 GeV.
- Un anillo de almacenamiento de muchos metros de diámetro por el que viajan electrones de alta energía.
- Y una serie de instalaciones con laboratorios experimentales en posición tangencial al último anillo.

Los electrones tienden a viajar en línea recta y, si son desviados, emiten un gran espectro de radiación electromagnética de alto brillo (Figura 8.9). Por este motivo, las líneas experimentales se sitúan de manera tangencial al anillo de almacenamiento, recogiendo la radiación liberada por los electrones en el giro. Dicha radiación se filtra mediante dispositivos ópticos en la línea experimental con la finalidad de obtener las energías necesarias para cada experimento de difracción. Por último, la radiación llega al equipo donde se sitúa el cristal y el detector (Figura 8.9). Las recogidas de datos de difracción utilizando radiación sincrotrón a lo largo de este trabajo se han llevado a cabo en las estaciones ALBA (Barcelona, España) y ESRF (Grenoble, Francia) en las líneas BL-13 XALOC e ID23-2 respectivamente.

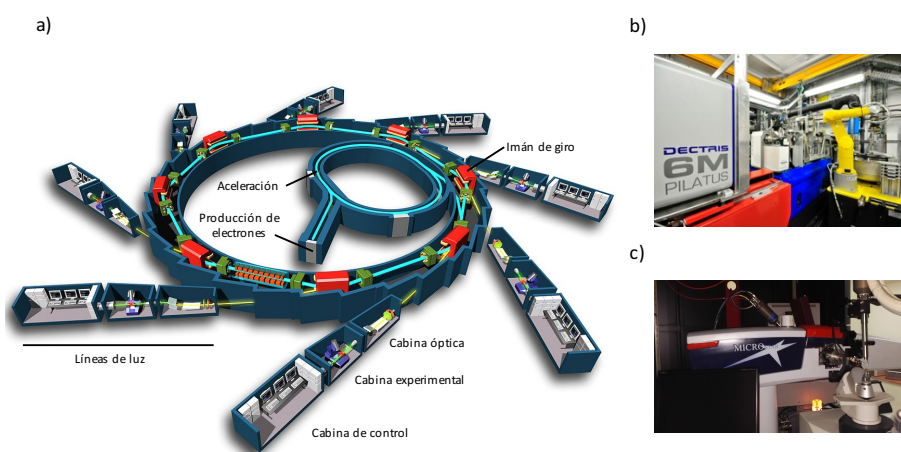


Figura 8.9. Fuentes de rayos X a) Esquema general del funcionamiento de una estación sincrotrón. b) Equipo para Cristalografía de macromoléculas en la línea de medida XALOC instalado en ALBA. c) Equipo de ánodo rotatorio *MicroStar 2,7 kW* instalado en el Instituto de Química Física Rocasolano (IQFR-CSIC). Imágenes a y c tomadas de: <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

En cuanto a los generadores de ánodo rotatorio, son equipos que contienen un ánodo móvil, lo que facilita la disipación del calor producido y permite el paso de corriente de más alto voltaje y, con ello, una mayor potencia de los rayos X que los obtenidos con un tubo convencional. Este tipo de generador se encuentra en el Instituto de Química Física Rocasolano (IQFR-CSIC), en el departamento de Cristalografía y Biología Estructural donde se dispone del modelo *MicroStar 2,7 kW* de Bruker. Pese a

que no se obtienen datos a tan alta resolución como los recogidos utilizando radiación sincrotrón, se usan muy frecuentemente para hacer un cribado inicial de cristales, distinguir entre cristales de sal o proteína, comprobar la máxima resolución, así como para probar distintos crioprotectores.

8.4.3 Detección de rayos X

Finalmente, los rayos X difractados por el cristal se registran en diferentes tipos de detectores que permiten registrar sus intensidades y digitalizarlas para su posterior procesamiento. Para la resolución estructural de las proteínas presentadas en este trabajo se han utilizado los detectores de matriz de píxeles que se basan en las tecnologías CCD, CMOS, HPAD y PILATUS. Estos tipos de detectores son los más utilizados en los sincrotrones actualmente. Contienen una capa de silicio muy sensible que absorbe los rayos X y los convierte en una señal eléctrica. El sistema de lectura es extremadamente eficiente y rápido, y genera además muy poco ruido de fondo (Figura 8.10).

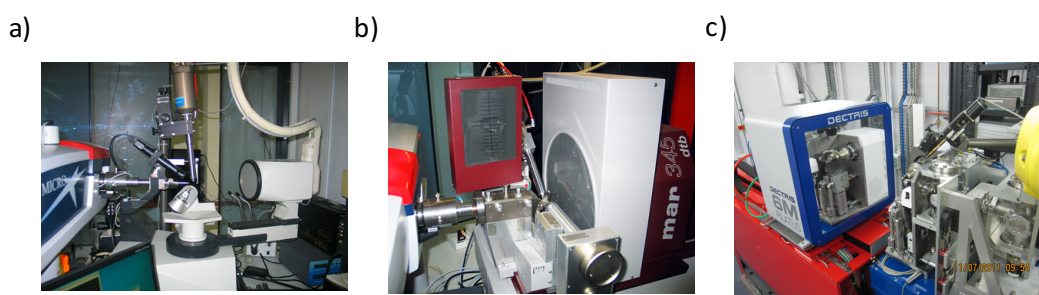


Figura 8.10. Sistemas de detección de rayos X a) Detector CCD y b) detector MAR345dtb, MarResearch ambos del Instituto de Química Física “Rocasolano” (IQFR-CSIC). c) Detector Pilatus 6M instalado en la línea XALOC del sincrotrón ALBA. Imágenes a y b tomadas de: <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

8.5 Procesamiento de datos de difracción de rayos X y resolución estructural

Los datos adquiridos durante el experimento de difracción han de procesarse de forma adecuada para que permitan la determinación de la estructura atómica de los cristales. La Figura 8.11 muestra las diferentes etapas que esto implica.

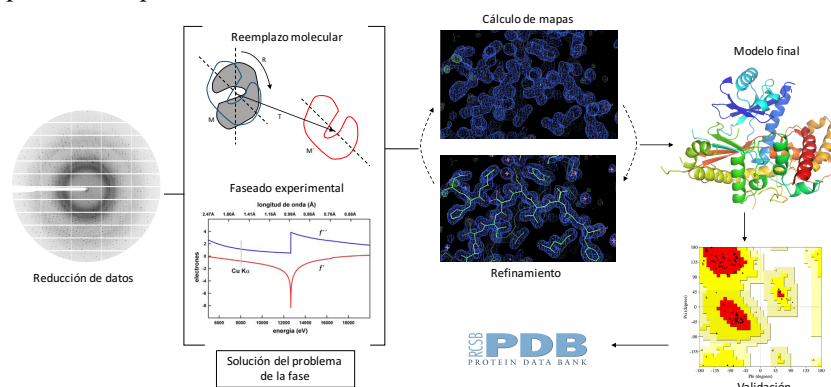


Figura 8.11. Etapas del proceso de resolución estructural a partir de datos de difracción de rayos X. En primer lugar, se analizan y procesan los datos, posteriormente se soluciona el problema de la fase (principalmente mediante

reemplazo molecular o faseado experimental) para después calcular los primeros mapas de densidad electrónica. Tras varios ciclos de refinamiento se obtiene un modelo que, si pasa la validación, estará listo para depositar en las bases de datos públicas (PDB).

8.5.1 Procesado y escalado de datos

El resultado del experimento de difracción es un conjunto de imágenes digitales en el que se encuentran registradas las intensidades difractadas por el cristal. Se puede dividir el procesado de los datos de difracción en varias fases: indexado, integración, escalado y, por último, el cálculo de los módulos de los factores de estructura.

Durante el indexado se analiza la geometría del espectro de difracción, lo que permite conocer ciertos parámetros del cristal como orientación, celdilla unidad y posible grupo espacial. Tras el indexado se realiza una etapa de integrado que consiste en estimar las intensidades (I_{hkl}) para cada reflexión hkl medida (los índices hkl definen un plano de átomos en el cristal). El proceso de escalado consiste en la corrección de las intensidades obtenidas debido a posibles defectos en la heterogeneidad del cristal o al daño por radiación, además hay que tener en cuenta parámetros como el ruido de fondo de la imagen, la mosaicidad o el decaimiento de la difracción durante el proceso de recogida. Por ejemplo, los datos tomados al inicio del experimento no presentan generalmente la misma calidad que los tomados al final del experimento. Durante este proceso, las reflexiones con los mismos índices hkl y las relacionadas por simetría cristalográfica son promediadas. Del proceso de escalado se obtienen unos parámetros que son empleados para evaluar la calidad general de los datos obtenidos. Algunos de estos parámetros son: los factores R_{merge}/R_{sym} , $I/\sigma(I)$, completitud, multiplicidad, $CC_{1/2}$ y CC^* .

- R_{merge} : factor estadístico de desacuerdo entre las intensidades de las reflexiones relacionadas por simetría dentro del patrón de difracción que evalúa la diferencia entre la intensidad de una reflexión medida N veces y su promedio. Un valor bajo en este factor indica un buen acuerdo entre las diferentes medidas de cada reflexión durante todo el experimento.
-
- R_{pim} : factor estadístico que tiene en cuenta en número total de reflexiones. Valores máximos de 0,3-0,4 para la última capa de resolución son aceptables.
- $I/\sigma(I)$: representa la relación de la intensidad con respecto al ruido de fondo. Son aceptables valores mayores de 1.5 para la última capa de resolución.
- Completitud: indica el porcentaje de espacio recíproco analizado. En la próxima sección se analiza en detalle.
- Multiplicidad: es el número de veces que se ha registrado una determinada reflexión. Cuanto mayor sea su valor, mejor será la estimación de errores asociados a la medida de la reflexión.
- $CC_{1/2}$ y CC^* : son dos parámetros basados en el cálculo del coeficiente de correlación de Pearson (CC). Estos parámetros consideran que, si se divide el conjunto de datos en dos, los errores en cada uno serían aleatorios.

Finalmente, considerando un balance de todos estos parámetros puede determinarse la resolución máxima de los datos de difracción. A mayor resolución, mejor definición de los mapas de densidad electrónica que se obtienen, lo que permite una mejor determinación estructural.

8.5.2 Problema de la fase y obtención de mapas de densidad electrónica

La transformación del espacio recíproco (espectro de difracción) al espacio real o cristalino (función de densidad electrónica) implica la resolución de una función denominada transformada de Fourier (Figura 8.12). Esta ecuación relaciona la densidad electrónica, $\rho_{x,y,z}$ para cada punto de celdilla elemental cristalina con el espacio recíproco, que está definido por los factores de estructura F_{hkl} . Estos factores de estructura (F_{hkl}) se definen como la resultante de la dispersión de rayos X por todos los átomos de la celdilla en una determinada dirección. Esta resultante es una onda y como tal, tiene un vector asociado que puede describirse por su módulo, $|F_{hkl}|$, y su fase ϕ_{hkl} . Los módulos representan las amplitudes de los haces difractados y se obtienen directamente a partir de las intensidades recogidas. Sin embargo, las fases ϕ_{hkl} no pueden determinarse experimentalmente y como se verá más adelante se requiere de distintos métodos para su cálculo.

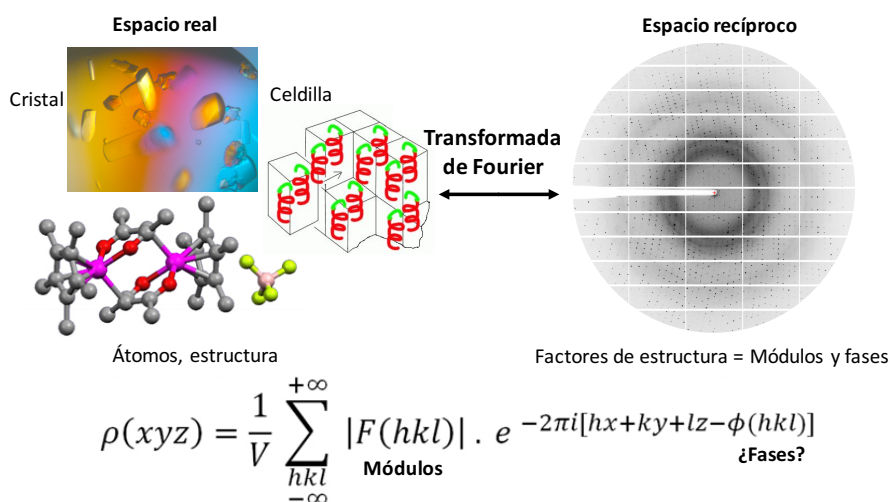


Figura 8.12. Problema de la fase. La relación entre el experimento de difracción (espacio recíproco) y la estructura de un cristal (espacio directo) está mediada por una transformada de Fourier, representada por la función de densidad electrónica: $\rho(xyz)$. Modificado de: <http://www.xtal.igfr.csic.es/Cristalografia/index.html>.

Por tanto, para calcular los mapas de densidad electrónica sobre los cuales se va a construir el modelo, es necesario conocer la fase ϕ_{hkl} asociada a cada haz difractado. Esto constituye el llamado “problema de la fase” en Cristalografía de proteínas. La información sobre cada fase puede recuperarse empleando diferentes aproximaciones. Las dos técnicas más utilizadas se describen brevemente a continuación:

8.5.2.1 Método de reemplazo molecular (MR)

Este método es uno de los más utilizados a día de hoy, debido a la relativa sencillez que supone llevarlo a cabo y gracias a la existencia de un gran número de proteínas cuya estructura ha sido previamente resuelta (Rossmann, 2001; Scapin, 2013). Cuando se conoce la estructura de una proteína con cierta identidad de secuencia respecto a la proteína en estudio, ésta puede emplearse como modelo para el cálculo de unas fases iniciales. Para ello se requiere que el modelo tenga una identidad de secuencia con la proteína de al menos un 25 – 30%. Este método se basa en el cálculo de una función de rotación (R) (orientación) y una función de traslación (T) (posición de celdilla) para el modelo, de forma que lo “sitúa” en el lugar donde se encuentra nuestra proteína. Una vez encontrada la mejor orientación y posición, se calculan las fases iniciales y se combinan con las amplitudes experimentales para calcular un mapa de densidad electrónica que permita trazar la cadena polipeptídica (Figura 8.13).

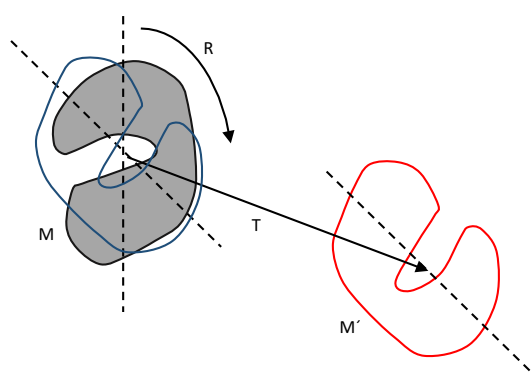


Figura 8.13. Esquema del método de reemplazo molecular. M representa la molécula de estructura conocida usada como modelo, que gira mediante la operación (R) y se traslada mediante la operación (T), para llevarla a la posición de la molécula desconocida (M'). Después debe evaluarse si la posición obtenida es correcta o si se debe seguir buscando otra orientación distinta. Modificado de: <http://www.xtal.iqfr.csic.es/Cristalografia/index.html>.

8.5.2.2 Método de dispersión anómala simple/múltiple (SAD/MAD)

Estos métodos se basan en la introducción ordenada en el cristal de átomos pesados que sean dispersores anómalos de los rayos X. Los átomos que se introducen en el cristal resuenan al ser irradiados a determinada longitud de onda, lo que genera una dispersión anómala detectable en el patrón de difracción. Las diferencias de intensidad detectadas producidas por este efecto se utilizan para estimar las fases iniciales. Si se utiliza una única longitud de onda durante la medida de los datos se habla de dispersión anómala simple (SAD, “*Single-wavelength Anomalous Dispersion*”) mientras que si se miden diferentes patrones a diferentes longitudes de onda se aplica el método de dispersión anómala múltiple (MAD, “*Multi-wavelength Anomalous Dispersion*”) (Hendrickson, 1991). La calidad de la señal anómala que se obtiene depende, entre otros factores, del número de dispersores anómalos unidos al cristal y de su capacidad dispersiva en el límite de absorción (Boggon y Shapiro, 2000). En este trabajo se han utilizado la técnica SAD como método para asignar fases al espectro de la proteína Nrd1. La técnica utilizada para introducir los átomos pesados en el cristal de proteína ha sido mediante sustitución de átomos de azufre de las metioninas por átomos de selenio, un fuerte dispersor anómalo, mediante protocolos de expresión de proteínas sustituidas con selenometionina. Se estima que es necesario un átomo de selenio por cada 100 aminoácidos para realizar Se-SAD (Strub et al., 2003).

Cada átomo dispersa la luz en función de su factor de dispersión atómico, f . En un experimento de tipo SAD o MAD, el factor de dispersión atómico (f) se divide en dos componentes, normal (f_0) y anómala (f' y f''), debido a la resonancia del átomo a determinada longitud de onda. Para maximizar estas diferencias anómalas en un experimento SAD, los datos de difracción de rayos X se miden a las longitudes de onda que se corresponde con la máxima capacidad dispersiva del átomo dispersor anómalo en cuestión. Para ello es necesario sintonizar la fuente de rayos X para que emita con la energía necesaria y, en ocasiones, esta se sitúa cerca de los límites de las líneas experimentales de radiación sincrotrón (habitualmente entre 5 y 15 keV).

En muchas ocasiones, las fases obtenidas no son lo suficientemente precisas como para permitir la construcción del modelo. En estos casos se emplean métodos de modificación de densidad para mejorar la calidad de los mapas obtenidos. Existen diferentes técnicas como el aplanado del solvente o el promediado por simetría no cristalográfica.

Existen otros métodos de resolución como las técnicas de reemplazo isomorfo/múltiple (SIR/MIR) o métodos directos (*ab initio*). Este último trata de obtener información de las fases directamente desde los módulos de los factores de estructura, usando relaciones de probabilidad.

8.5.3 Construcción del modelo estructural y refinado

Una vez asignada a cada reflexión su fase, se puede calcular la función de densidad electrónica. La representación gráfica de esta función origina el mapa de densidad electrónica ($2F_o - F_c$), así como el mapa de densidad electrónica diferencia ($F_o - F_c$) donde los picos de densidad positiva revelan que el modelo está incompleto, mientras que los negativos muestran posiciones incorrectas del mismo. Mediante el manejo de programas gráficos de modelado (*Coot*) se pueden ir introduciendo cambios en el modelo, ajuntándolos al mapa de densidad electrónica. Una vez construido el modelo, se lleva a cabo un refinamiento matemático con el objetivo de minimizar las diferencias entre los factores de estructura calculados del modelo (F_c) con los factores de estructura experimentales (F_o). Con cada ciclo de refinamiento se calculan mapas de densidad electrónica de mejor calidad que permiten ir completando el modelo. Existen distintos tipos de refinamiento y los parámetros que se refinan son las coordenadas atómicas, los factores de temperaturas que indican la agitación térmica o movilidad de cada átomo y el grado de ocupación de un átomo en una determinada posición.

Para asegurar que los sucesivos ciclos de refinamiento y posterior construcción del modelo conducen hacia un modelo preciso, existen una serie de parámetros de calidad. Así, el factor de desacuerdo R_{Work} estima la calidad general del modelo.

$$R = \frac{\sum ||F_o| - |F_c||}{\sum ||F_o|}$$

Este valor es dependiente de la resolución y va disminuyendo con los progresivos ciclos de refinamiento e indica las diferencias entre los factores de estructura observados ($|F^{obs}|$) y los calculados ($|F^{calc}|$) a partir del modelo. Existe un segundo parámetro R_{Free} , se calcula con un 5-10% de las reflexiones que son excluidas del refinamiento. Con este segundo parámetro se consigue una validación cruzada que

elimina sesgos en el procedimiento matemático. El refinamiento se da por finalizado cuando el modelo estructural se ajuste al mapa de densidad electrónica con la mayor precisión posible.

8.5.4 Validación del modelo

La última fase del proceso de determinación estructural consiste en la validación del modelo obtenido. Mediante diferentes parámetros se comprueba que el modelo cumpla ciertas restricciones geométricas (ángulos y distancias de enlace entre átomos) y restricciones estereoquímicas (ángulos de torsión de la cadena polipeptídica) que se evalúan mediante el diagrama de Ramachandran. También se controla que no existan distancias entre átomos por encima o debajo de los parámetros permitidos (Figura 8.14). El resultado final proporciona las coordenadas atómicas del modelo y los correspondientes factores de estructura. Tanto las coordenadas como los factores de estructura son depositados en el Banco de Datos de Proteínas (PDB, *Protein Data Bank*) donde pasan un minucioso proceso de evaluación y validación. Finalmente, se les asigna un código numérico que identifica la estructura de la proteína (código PDB) y se pone a disposición del dominio público.

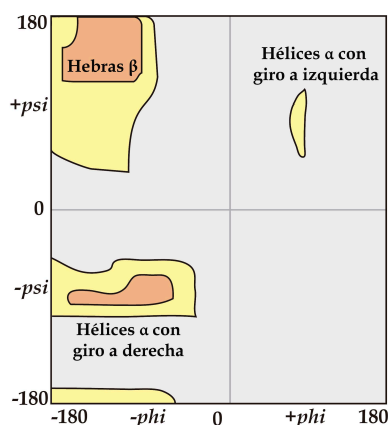


Figura 8.14. Diagrama de Ramachandran. En el diagrama de Ramachandran se puede visualizar todas las combinaciones posibles de ángulos de torsión de los enlaces peptídicos que contribuyen a la conformación de la estructura de proteínas.

